Expression of *Plasmodium falciparum var* genes in naturally infected children from Tanzania

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch – Naturwissenschaftlichen Fakultät
der Universität Basel

von

Joseph Paschal MUGASA
aus
Morogoro, Tanzania

Basel, 2008

Genehmigt von der Philosophisch-Naturwissenchaftlichen Fakultät der Universität Basel auf Antrag von

Prof. Dr. Hans-Peter Beck, Prof. Dr. Till Voss, Prof. Dr. Thomas Seebeck und Prof. Dr. Marcel Tanner

Basel, 11 April 2008

Prof. Dr. Hans-Peter Hauri

Dekan

Dedication

To my beloved parents

Paschal Thomas Mugasa & Paulina Francis Sizya

To my son ***Bryant Joseph Mugasa***

……..…brightest light in the darkest night……

Table of contents

Acknowledgements

This thesis was part of collaboration between the Ifakara Health Research and Development Centre (IHRDC) and Saint Francis Designated District Hospital (StFDDH) in Tanzania and Swiss Tropical Institute (STI). Thus many people in different disciplines contributed to the success of this work and to them all I extend my sincere appreciation. I say "*asanteni sana*" for all your efforts to make this work a success.

The writing of a PhD thesis is a long and arduous task. Only with the support of others can anyone survive the ordeal. My first and foremost gratitude goes to Prof Hans-Peter Beck of the Swiss Tropical Institute for mentoring this work. He provided me with the finest scientific counseling and guidance that I could ever wish for. I thank him accommodating and creating a working environment in his group. I appreciate very much his open mindedness for any idea and question and his uncomplicated way of handling complex issues. I remember we took a walk in the cold to Biozentrum looking for help in Bioinformatics which resulted in recruitment of Dr Weihog Qi at STI, Thank you! Weihong later provided invaluable help in sequence handling and analysis, running the simulation model was very fascinating…!! Definitely without her the smooth progress of my work would have been hindered.

My deepest thanks go to parents and guardians of children who willingly participated in this study thanks a lot! Without them it was not possible to carry out this study.

I wish to thank Prof. Dr. Till Voss, Prof. Dr. Thomas Seebeck and Prof. Dr. Marcel Tanner for being part of my thesis committee and for their support. I understand it is a demanding task in terms of time required in processing a thesis and report writing, I highly appreciate the effort.

Thanks to Dr Mirjam Kaestli for assisting in setting up the case control study at Ifakara and supervising the initial recruitment of patients and processing the samples at IHRDC laboratory. At STI Mirjam Käestli was also an instrumental in familiarizing me into this world of *var* genes!! Thanks a lot

In Ifakara, my deepest appreciation goes to the director of Ifakara Health Research and Development Centre, Dr Hassan Mshinda for his courageous, motivation and moral support on this challenging assignment. I extend my sincere appreciation to the "Molgene group" in Ifakara colleagues Dr Kefas Mugittu (head of Biomedical Unit), Boniface Jullu, Deborah Sumari, Siana Nkya, Valeriana Mayagaya, Cecil Lupala, Allen Malisa, Erasto Mbugi, Boniface Idindili, Charles Maswi, Eugen

Acknowledgements

Shirima Technicians: John Wigayi, Mzee Kobero, Tarsis Pius, Athumani Mtandaguo, Fidelist Mbena, Augustine Mahundi, Ben Liseki, Administrative and supporting staff at IHRDC: Mzee Urassa, David Matondo, Mama Eliza Yegella, Matilda, Idda Ruzige, Aveline Magoha, Bangaseka, Ally Mkesa, Kadeghe Aloys, Nicas Makinda, Sikana Lusekelo, Mashaka Lumuli, Grace Temu I thank you all for your hard working spirit which has made life easier. To all the staff at StFDDH working at pediatric ward thanks for assisting with clinical definition of cases and recruiting patients: Dr Bunini Manyilizu, Mrs Kasiga I, Sr Philipina Hongoa and all the nurses ….*Jamani asanteni sana na Mungu awabariki …*

In Basel, special thanks to Christine Mensch, Christine Walliser, and Magrit Sloui for the administrative support. To my friends who were great inspiration during my stay in Switzerland, Mirjam Brown, Martin Zumstein, Bonna Muller, Shinji Okitsu (uncle), Marco Tamborini, Daniella, Valentin, JP, Naomi Maina, Dorothy, Lucy Ocholla, Hamis Malebo, Rashid Ahmed, Kafura W, Mwanyika H, Susan Rumisha, Manfred Chaponda and Mr and Mrs Carstein. In Molecular Parasitology and Infection Biology (MPI): Selina Bopp, Christian Flueck (Mapu…), Nsanzabana, Caroline, Matthias Rottman, Silyvia Steiger, Dania Muller, Cornelia Spycher, Sebastian Rusch, Jutta Marfurt, Nicole Falk, Kathrin Witmer, Anouk Muller, Martin Maire, Dieter, Igor and Ingrid Felger *Danke schöne!!*

With deep gratitude, I acknowledge my parents and members of our family, dad Paschal Mugasa and mum Pauline Sizya for their profound influence on my life and on my thinking, you're the rocks upon which is the foundation of my achievements, and all of my education are laid, thanks for your loving memory guidance and trust. My brothers and sisters, Francis, James, Alberto, Dr Aveline, Felician, and Glory thanks for your courage and prayers, and lastly to my sweetheart Dr Nelly Iteba for her constant encouragement, unbelievable kindness and abundant trust.

Most importantly, I acknowledge and thank God for the blessings, insight, and support I have felt throughout this project. For me, he is the source of all principles that bring joy and success in life.

This work was financially supported by Optimus foundation and Swiss National Science Foundation.

Summary

Plasmodium falciparum is the most pathogenic malarial parasite and a major cause of morbidity and mortality among young children in sub-Saharan Africa. The virulence of *P. falciparum* has been linked to its expression of variant surface antigens (VSAs) on the surface of infected red blood cells. These VSAs subvert acquisition of protective immunity and mediate cytoadherence of infected erythrocytes to the microvasculature lining of various endothelial cell receptors. It causes sequestration of infected erythrocytes in post capillary venules of the vital organs such as the brain or placenta. Cytoadherence causes retention and accumulation of the infected erythrocytes to endothelial membranes of deep post-venous capillaries leading to occlusion of micro-vessels. This result in obstruction of free blood flow with serious pathological consequences associated with severe malaria. Sequestration facilitates parasite multiplication and enables the parasites to avoid the passage of infected erythrocytes through the spleen, where deformed erythrocytes are removed from blood circulations. This cytoadherence is mediated by *P. falciparum* erythrocyte membrane protein 1 (PfEMP1). PfEMP1 is a VSA family encoded by ~ 60 highly polymorphic *var* genes per haploid genome, expressed on the surface of infected red blood cells. PfEMP1 is expressed in a mutually exclusive manner, and switching the expression creates extensive antigenic variation and the potential for multiple adhesion profile. Antigenic variation is a strategy employed by *P. falciparum* to avoid antibody-mediated destruction by alternating expression of individual *var* genes each of which encodes an antigenically distinct form of PfEMP1. Sequence analysis of the *var* gene repertoire of the 3D7 clone revealed genetic structuring in which *var* genes fall into 3 distinct groups (A, B, and C) and two intermediate groups (B/A and B/C) based on chromosomal location, gene orientation and the 5' flanking sequences. It has been postulated that this genetic organization helps to restrict recombination within a specific group of genes and leads to their structural and functional specialization for binding to different endothelial receptors.

Summary

The sequences of *var* genes vary substantially within and between the parasites genome. This has been clearly indicated by the fact that there is minimal overlap in the *var* gene repertoire between isolates due to high inter-genic and intra-genic recombination within the *var* gene family. Despite the complex nature of this molecule, the *var* gene still remains the best defined factor contributing to malaria pathogenesis. Different research groups have attempted to define the repertoire of *var* gene from different isolates, and reported vast global *var* gene diversity. Only a tip of iceberg of the *var* genes diversity is currently in view. The big challenge to date is to understand how the *var* gene diversity and selection pressure influence malaria pathogenesis in order to device a control strategy based on interference with PfEMP1 expression.

Clinical and sero-epidemiological studies have suggested that severe disease is attributed by the parasite expressing a restricted and antigenically conserved subset of VSAs which are frequently recognized by sera from semi-immune individuals, proposing that expression of a particular VSA may be associated with disease manifestation. Pregnancy associated malaria (PAM) is well understood and has often been linked with the expression of a *var* gene called *var2csa* which is unusually conserved across parasite isolates and binds a low sulfated form of chondroitin sulfate A (CSA) in the placenta. Different studies have attempted to link a particular *var* gene expression with a disease phenotype. It is becoming evident that *var* group A and B/A are involved in severe childhood malaria. Protective immunity to severe malaria develops earlier in childhood after only few severe episodes pointing to a relatively conserved target antigen. This phenomenon makes it theoretically possible to protect non immune children against severe and complicated malaria by accelerating acquisition of PfEMP1 specific immunity.

Summary

Given the proposed importance of immunity to PfEMP1 in protection against malaria, it is essential that we gain a better understanding of *var* gene expression during infection. Despite substantial contribution of *var* genes to malaria pathogenesis and parasites survival, few studies on *var* gene transcription during natural infections have been carried out in field isolates. This is mainly attributed to technical difficulties, and the complexity and immense diversity interfering with most study design.

For this thesis, two studies on *var* gene expression in naturally infected children with severe *P. falciparum* malaria from Tanzania were conducted. In the first study, the transcription levels of *var* gene groups were compared in children with severe, uncomplicated and asymptomatic malaria by using quantitative real-time PCR. Transcripts of *var* group A and B genes were up-regulated in children with severe malaria compared to patients with uncomplicated malaria. In general, the transcript abundances of *var* group A and B genes were higher for children with clinical malaria than for children with asymptomatic infections. *var* group C was not linked with any disease phenotype.

In the second study, the genetic diversity of expressed *P. falciparum var* genes in children with severe malaria from Tanzania was determined. The *var* transcripts isolated from children with severe malaria (Blantyre score ≤ 3) were compared with isolates from children with asymptomatic malaria. Diversity patterns of dominant full-length *var* transcripts were determined by isolation of mRNA followed by magnetic bead capture through an ATS-anchor and reverse-transcription into *var* cDNA. The different PCR amplified expressed sequence tags were cloned and sequenced. Large sequence diversity of the amplified *var* DBL-1α and the 5′ non-coding regions was observed and minimal overlapping was evident among the isolates providing strong evidence that the transcribed *var* gene repertoire is immense. *var* DBL-1α sequences isolated from AM were more diverse

with more singletons (P<0.05) compared with DBL-1α sequences from SM. Unique *var* sequences that were exclusively expressed with *P. falciparum* isolated from children with SM were found. Despite the fact that *var* gene diversity is unlimited, transcripts from SM isolates were more restricted, supporting the hypothesis that certain PfEMP1 repertoires are involved in triggering severe infections.

Zusammenfassung

Plasmodium falciparum ist der Hauptauslöser von Malariapathologie, Morbidität und Mortalität bei kleinen Kindern in Afrika südlich der Sahara. Virulenz von *P. falciparum* ist abhängig von der Expression der ‚variant surface antigens' (VSAs, variable Oberflächen-Antigene) auf der Oberfläche der infizierten roten Blutzellen. Diese VSAs verhindern die Entwicklung einer schützenden Immunantwort und führen zu Zytoadhärenz der infizierten Erythrozyten an Rezeptoren des Endothels. Dies ermöglicht die Sequestrierung der infizierten Erythrozyten in den Blutkapillaren wichtiger Organe, wie z.B. im Gehirn oder in der Plazenta. Zytoadhärenz führt zur Rückhaltung und Anhäufung infizierter Blutkörperchen und damit zur Blockierung der Mikro-Kapillaren. Dies wiederum verhindert den Blutfluss und führt zu schweren pathologischen Komplikationen, wie z.B. zu zerebraler Malaria. Sequestrierung ermöglicht aber auch die Parasitenmultiplikation und ermöglicht es dem Parasiten nicht durch die Milz zu passieren, in welcher deformierte Erythrozyten aus der Zirkulation eliminiert werden. Diese Zytoadhärenz wird durch ein Parasiten-Protein vermittelt, dem *P. falciparum* Erythrocyte Membrane Protein 1 (PfEMP1, Erythrozyten Membran-Protein 1). PfEMP1 gehört zur Familie der VSAs, und etwa 60 hoch-polymorphe *var* Gene kodieren für dieses Protein im haploiden Genom. Nur ein PfEMP1, exprimiert von einem einzelnen *var* Gen, ist auf der Oberfläche der infizierten roten Blutkörperchen exprimiert und durch Umschalten der Expression wird eine ausgedehnte Antigenvariation generiert mit der Möglichkeit, an verschiedenste Rezeptoren zu binden. Antigenvariation ist eine Strategie von *P. falciparum* der Antikörper-abhängigen Zerstörung, durch Umschalten einzelner *var* Gene, zu entgehen. Sequenzanalysen des *var* Genrepertoires von 3D7 zeigte eine genetische Struktur, welche 3 distinkte Gruppen (A, B, und C), sowie zwei Zwischengruppen (B/A und B/C), definiert. Die Gruppierung basiert auf der chromosomalen Lage,

der Genorientierung, sowie der 5' flankierenden Sequenzen. Es wurde postuliert, dass diese genetische Struktur dazu beiträgt, Rekombination nur auf eine bestimmte *var* Gengruppe zu begrenzen, und somit zur strukturellen und funktionellen Spezialisierung der Bindung an verschiedene Endothel-Rezeptoren beigetragen hat.

var Gensequenzen variieren substantiell innerhalb und zwischen verschiedenen Parasiten-Genomen. Dies wird besonders sichtbar bei der minimalen Überlappung des *var* Genrepertoires verschiedener Parasiten-Isolate, welche vermutlich durch die hohe inter- und intra-genische Rekombnation innerhalb der *var* Genfamilie bedingt ist. Trotz ihrer Komplexität sind die *var* Gene die best untersuchten Pathologiefaktoren bei Malaria. Verschiedene Forschungsgruppen haben versucht, das *var* Genrepertoire verschiedener Parasiten-Isolate zu definieren und berichten von einer riesigen globalen Diversität. Offensichtlich repräsentiert dies aber nur erst die Spitze des Eisberges. Und eine grosse Herausforderung wird es sein, zu verstehen, wie *var* Gendiversität und Selektionsdruck Malaria-Pathogenese beeinflusst, um dann entsprechende Interventionsstrategien zu entwickeln, die darauf beruhen, die Expression von PfEMP1 zu unterdrücken.

Klinische und sero-epidemiologische Studien deuten darauf hin, dass schwere Malaria durch Parasiten ausgelöst wird, welche eine beschränkte und antigenisch konservierte Untergruppe von *var* Genen exprimieren. Diese VSAs werden häufig von Seren von semi-immunen Menschen erkannt, was darauf hinweist, dass die Expression eines bestimmten VSAs mit der Ausprägung der Krankheit assoziiert ist. Schwangerschafts-assoziierte Malaria (PAM, pregnancy associated malaria) wird inzwischen besser verstanden, und sie wurde häufig mit der Expression eines spezifischen *var* Gens, des *var2csa*, in Verbindung gebracht. Dieses *var* Gen

ist aussergewöhnlich stark konserviert zwischen verschiedenen Parasiten-Isolaten und bindet Chondroitinsulfat A (CSA) in der Plazenta. Verschiedene Studien haben versucht, die Expression individueller *var* Gene mit einem Krankheitsphänotyp in Verbindung zu bringen. Es gibt vermehrt Hinweise darauf, dass die *var* Gruppe A und B/A bei schwerer Kindermalaria involviert sein könnte. Schützende Immunität gegen schwere Malaria entsteht früher in der Kindheit, bereits nach wenigen schwereren Episoden, was wiederum auf wenige relativ konservierte Ziel-Antigene hindeutet. Diese Tatsache würde es theoretisch möglich machen, nicht-immune Kinder gegen schwere Malaria oder gegen Malaria mit Komplikationen zu schützen, in dem man die Entwicklung der spezifischen Immunität gegen PfEMP1 unterstützt.

Wegen der Wichtigkeit der Immunität gegen PfEMP1 beim Schutz gegen Malaria, ist es wichtig, besser zu verstehen, welche Regeln der Expression von *var* Genen in natürlichen Infektionen zu Grunde liegen. Trotz der wichtigen Rolle, die *var* Gene in der Pathogenese und beim Überleben der Parasiten spielen, gibt es nur wenige Studien, die die *var* Genexpression in natürlichen Infektionen in Feld-Isolaten angeschaut haben. Der Grund hierfür liegt in technischen Schwierigkeiten und an der immensen Komplexität und Diversität, welche die Durchführung der meisten Studien limitiert.

Im Rahmen dieser Doktorarbeit wurden zwei Studien über *var* Genexpression in natürlichen Infektionen bei Kindern mit schwerer Malaria in Tanzania durchgeführt. In der ersten Studie wurde die Expression der verschiedenen *var* Gengruppen in Kindern mit schwerer Malaria, in Kindern mit leichter klinischer Malaria, und in Kindern mit asymptomatischer Malaria mittels quantitativer ‚real-time' PCR verglichen. *var* Transkripte der *var* Gruppe A und B waren in Kindern mit schwerer Malaria hoch reguliert, verglichen mit Kindern mit leichter

klinischer Malaria. Insgesamt waren die *var* Gene der Gruppe A und B in Kindern mit klinischen Symptomen höher transkribiert, verglichen mit Kindern mit asymptomatischer Malaria. Die Expression der *var* Gengruppe C war mit keinem Krankheitstyp assoziiert.

In der zweiten Studie wurde die genetische Diversität der exprimierten *var* Gene in Kindern aus Tanzania mit schwerer Malaria determiniert. Die *var* Transkripte wurden von Kindern mit schwerer Malaria (Blantyre score ≤ 3) isoliert und mit Transkripten aus Kindern mit asymptomatischer Malaria verglichen. Das Diversitätsmuster der dominanten ‚full-length' *var* Transkripte wurde durch Isolation von mRNA über Bindung an Magnetkügelchen mit der ATS-Domäne und anschliessender reverser Transkription in cDNA bestimmt. Die verschiedenen PCR amplifizierten exprimierten Fragmente wurden kloniert und sequenziert. Dadurch wurde in den amplifizierten DBL1α und den 5' nicht-kodierenden Regionen eine grosse Sequenzdiversität sichtbar. Zwischen den Sequenzen einzelner Isolate bestand fast keine Überlappung, was auf ein unlimitiertes Repertoire der exprimierten *var* Gene hinweist. Die DBL1α *var* Sequenzen aus asymptomatischen Malariafällen war diverser, mit einer grösseren Anzahl von Einzelsequenzen als die DBL1α Domänen aus schweren Malariafällen ($p<0.05$). Es wurden spezifische *var* Sequenzen, die nur bei schwerer Malaria auftraten, gefunden. Obwohl die *var* Gendiversität so gross ist, war die Diversität der Transkripte aus schweren Malariafällen eher beschränkt, was die Hypothese unterstützt, dass nur eine bestimmte Anzahl spezifischer *var* Gene bei der Auslösung von schwerer Malaria beteiligt ist.

Abbreviations

AM	Asymptomatic Malaria
ATS	Acidic Terminal Segment
CD 36	Clusters Determinant 36
CIDR	Cystein-rich InterDomain Regions
CR1	Compliment Receptor 1
CSA	Chondroitin sulfate A
DBL	Duffy Binding-like
DBP	Duffy binding proteins
DNA	Deoxyribonucleic Acid
dNTP	deoxyribonucleoside triphosphate
EBA	Erythrocyte Binding Antigen
HA	Hyaluronic Acid
ICAM-1	Intracellular adhesion molecule 1
IHRDC	Ifakara Health Research and Development Centre
MOI	Multiplicity of infections
mRNA	Messenger Ribonucleic Acid
MSP2	Merozoite Surface Protein 2
NTS	N-terminal segment
PAM	Pregnancy Associated Malaria
PCR	Polymerase Chain Reaction
PECAM	Platelet endothelial cell adhesion molecule 1
PfEMP1	*Plasmodium falciparum* membrane protein 1
pRBC	parasitized Red Blood Cells
qRT-PCR	quantitative Real-Time Reverse Transcription PCR
RDT	Rapid Diagnostic Test
RFLP	Restriction Fragment Length Polymorphism
RIFIN	Repetitive Interspersed Family
RNA	Ribonucleic Acid
SM	Severe Malaria
STEVOR	Subtelomeric Variable Open Reading Frame Family
StFDDH	St Francis Designated District Hospital
STs	Sequence Types
TM	Transmembrane Domain
TSP	Thrombospodin
UM	Uncomplicated Malaria
VCAM-1	Vascular Cell Adhesion Molecule-1
VSA	Variant Surface Antigen
WHO	World Health Organization

Chapter One

Introduction

Chapter One

1.0 Introduction

Plasmodium falciparum malaria, besides tuberculosis and HIV, is a major global health problem. It accounts for more than 500 million clinical cases per year, mainly in children < 5 years and primigravid women in Sub-Saharan Africa (Breman 2001). Every 40 seconds a child dies of malaria, resulting in a daily loss of more than 2000 young lives worldwide. In addition to reducing quality of life, malaria also imposes a heavy economic burden on developing countries (Malaney et al 2004, Sachs & Malaney 2002). Despite extensive research efforts, no effective vaccine capable of conferring an adequate level of immunity has been developed to date. Furthermore, rapid emerging drug resistance in natural parasite populations and the arising of insecticide-resistant mosquitoes, highlights the need for new intervention strategies that are both effective in the treatment and prevention of the disease.

Individuals living in areas of high *P. falciparum* transmission acquire protective immunity to severe malaria during early childhood after only a few symptomatic infections. However, they remain susceptible to uncomplicated disease and asymptomatic infection into adulthood. Thus, sterile immunity that prevents infection may never develop, but significant antidiseases immunity is acquired relatively rapidly.

Clinical manifestations of *falciparum* malaria differ markedly from infection to infection, although disease symptoms often remain uncomplicated. However, in some cases severe complications such as cerebral malaria, severe anaemia or respiratory distress develop (Marsh et al 1995). The reasons why certain children develop life-threatening complications while others are able to tolerate very high parasite burdens without severe clinical symptoms remain unclear (Miller et al 2002).

Severe malaria has previously been associated with expression of a restricted and antigenically conserved subset of variant erythrocyte antigens (Bull et al 2000, Nielsen et al 2002). This suggests that expression of certain surface molecules may be associated with specific disease manifestations. Malaria parasites causing clinical disease in semi immune patients express variant surface antigens (VSA) that correspond to the 'holes' in the VSA antibodies repertoire. The parasites expressing VSA, to which there is no pre-existing acquired immunity can multiply in a substantial way, leading to clinical diseases. Acquisition of protective immunity involves sequential closure of these holes. Immunity to severe malaria is relatively quick to develop after a few episodes (Gupta et al 1999), indicating the existence of antigenic homogeneity in parasites causing severe disease.

The best characterized VSA are the *var* genes encoded *Plasmodium falciparum* Erythrocyte Membrane Protein 1 (PfEMP 1). There are about 60 *var* genes per haploid genome (Gardner et al 2002). PfEMP 1 is expressed in a mutually exclusive manner at the surface of infected erythrocyte. It mediates parasites to bind to host endothelium and other host cells. Sequestration in microvessels allows parasitized blood cells (pRBCs) to avoid clearance from blood stream by the spleen. Switching of the *var* gene expression allows the parasites to modify the antigenic and functional properties of parasitized erythrocytes, thereby evading immunity and affecting the outcome of infection.

The introduction chapter of this thesis is composed of three parts. The first part focuses on the parasites biology; the second part highlights the clinical and molecular aspects of severe malaria whilst the third part explains the role of PfEMP1 in malaria pathogenesis.

1.1 *Plasmodium* and Malaria

Malaria is caused by an infection with a protozoan parasite of the genus *Plasmodium,* which is transmitted through the bite of an infected female *Anopheles* mosquito. Of the approximately 400 species of *Anopheles* throughout the world, about 60 are malaria vectors under natural conditions, 30 of which are of major importance. Malaria parasites are eukaryotic single-celled microorganisms that belong to the genus *Plasmodium*. *Plasmodia* are members of the phylum *Apicomplexa*, characterized by the presence of an apical complex, which contains an apicoplast, a polar ring organizing the microtubules, the vesicles called micronemes, rhoptries and dense granules. The genus *Plasmodium* contains more than 100 species of which four infect humans: *Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale* and *P. malariae*. Recently there has been an outbreak of *P. knowlesia* in Borneo, Malaysia, and reports of human infections in Asia are numerous (Cox-Singh et al 2008). Whether human to human transmission occurs is yet unknown. These major species differ morphologically, immunologically, in their geographically distribution, in their relapse patterns and in their drug responses. Of the five species of malaria parasites, *P. falciparum* is the most virulent and responsible for severe pathogenesis such as cerebral malaria (coma), severe anaemia, renal failure, respiratory distress, metabolic acidosis, hypoglycemia and lung oedema. *P. falciparum* is the principle cause of malaria death in young children and pregnant mothers in endemic countries (Breman et al 2001). The least common malaria parasite is *P. ovale*, which is found throughout the world, natural distribution is in sub-Saharan Africa and Islands of Western Pacific. *P. ovale* has also been reported in South East Asia and South Pacific (Collins & Jeffery 2005). *P. malariae* occurs at low frequency in a patchy distribution worldwide. The most widespread malaria parasite is *P. vivax* but infections with this species are rarely fatal, although recent reports have linked it

with cerebral malaria (Kochar et al 2007). How *P. vivax* causes cerebral malaria remains a mystery. *P. falciparum* and *P. vivax* can both cause severe blood loss (anemia), mild anemia is more common in *P. vivax* infections, whereas severe anemia in *P. falciparum* malaria is a major cause of death in Africa. A characteristic feature of *P. falciparum* malaria is the ability of the parasite to invade red blood cells (RBCs) of all ages causing very high parasitemia, high multiplication rates (approximately 24 merozoites as compared to 8-10 merozoites in *P. vivax*) and enhanced growth, as well as the capacity to adhere to host endothelium (cytoadherence) and to non-infected RBCs (rosetting). These binding events eventually lead to the occlusion of the microvasculature in various tissues and organs, such as the brain in cerebral malaria (Miller et al 2002) hence contributing directly to the pathogenesis of severe malaria disease. *P. ovale* and *P. vivax* have dormant liver stages named hypnozoites that may remain in this organ for a period ranging from weeks to many years before the onset of a new round of pre-erythrocytic schizogony, resulting in relapses of malaria infection. In some cases *P. malariae* can produce long-lasting blood-stage infections, which, if left untreated, can persist asymptomatically in the human host for periods extending into several decades.

Mortality is not only the problem with malaria since morbidity in endemic countries leads to major socio-economic losses. In Tanzania for instance, malaria is the leading cause of out-patient and in-patient health service attendance for all ages, and is the leading cause of death in both children and adults in all regions of the country. Malaria is believed to be directly or indirectly responsible for about 16 million annual malaria episodes and 100,000 to 125,000 annual deaths in Tanzania of whom 70–80,000 are children under-fives (Ministry of Health Government of Tanzania 2003).

1.2 Life cycle of *Plasmodium falciparum*

The life cycle of all *Plasmodium* species is extremely complex and requires specialized protein expression for survival in both the invertebrate and vertebrate hosts. These proteins are required for both intracellular and extracellular survival, for the invasion of a variety of cell types and for the evasion of host immune responses. The life cycle of the malaria parasite is shown in (Figure 1) and can be divided into three consecutive phases of multiplication: Two phases of schizogony (asexual multiplication) in the vertebrate host, first in hepatocytes then in RBCs and one phase of sporogony (sexual multiplication) in the mosquito. Infection in vertebrate begins through the bite of an infected female *Anopheline* mosquito. Sporozoites released from the salivary glands of the mosquito enter the bloodstream, quickly reach the liver and penetrate the liver cells (hepatocytes) where they remain for 5-16 days for *P. falciparum* and undergo asexual replication known as exo-erythrocytic schizogony to form hepatic schizonts (Kappe et al 2004). The mechanism of targeting and invading the hepatocytes is not yet well understood, but studies have shown that sporozoite migration through several hepatocytes in the mammalian host is essential for completion of the life cycle (Mota et al 2001). The receptors on sporozoites responsible for hepatocyte invasion are mainly the thrombospondin domains on the circumsporozoite protein and on thrombospondin-related adhesive protein. These domains specifically bind to heparan sulfate proteoglycans on the hepatocytes (Frevert et al 1993). Each schizont gives rise to up to 10,000s merozoites inside the hepatocyte and each merozoite can invade a RBC on release from the liver. In the RBC they multiply within 48 hours giving rise to approximately 24 merozoites for *P. falciparum*, which are released and again invade RBCs thereby maintaining the erythrocytic cycle. The clinical manifestations of malaria, fever and chills are associated with the rupture of the infected erythrocyte. Not all of the merozoites divide into schizonts,

as some differentiate into sexual forms, male and female gametocytes. These gametocytes are taken up by a female anopheles mosquito during a blood meal. Within the mosquito midgut, the male gametocyte undergoes a rapid nuclear division, producing 8 flagellated microgametes which fertilize the female macrogamete to form zygotes. The zygotes formed by this fertilization develop into motile ookinetes, which invade and traverse the midgut epithelium. Diploid ookinetes undergo meiosis and, on reaching the basal side of the midgut, transform into oocysts, thereby undergoing several round of meiosis as they mature. Each oocyst releases thousands of haploid sporozoites into the mosquito hemocoel, from where they are transported through the hemolymph and invade the mosquito salivary glands. Sporozoites are finally transmitted to a new vertebrate host during an infective bite and the *Plasmodium* life cycle begins again reviewed by Whitten et al (2006).

The life cycle of *Plasmodium* consists of four invasive stages (a) the ookinetes traversing the intestinal cells in the mosquito (b) the sporozoites infecting the mosquito salivary glands (c) the vertebrate hepatocytes and (d) theamemaliza merozoites infecting the vertebrates' erythrocytes. The sporozoites and the hepatic stages are called the pre-erythrocytic stages. The hepatic stage is asymptomatic in humans and takes approximately 5-16 days in the case of *P. falciparum*. Clinical symptoms which can be severe are solely due to the erythrocyte stages. Almost all antimalarial drugs except primaquine are directed against this stage (Fidock et al 2004).

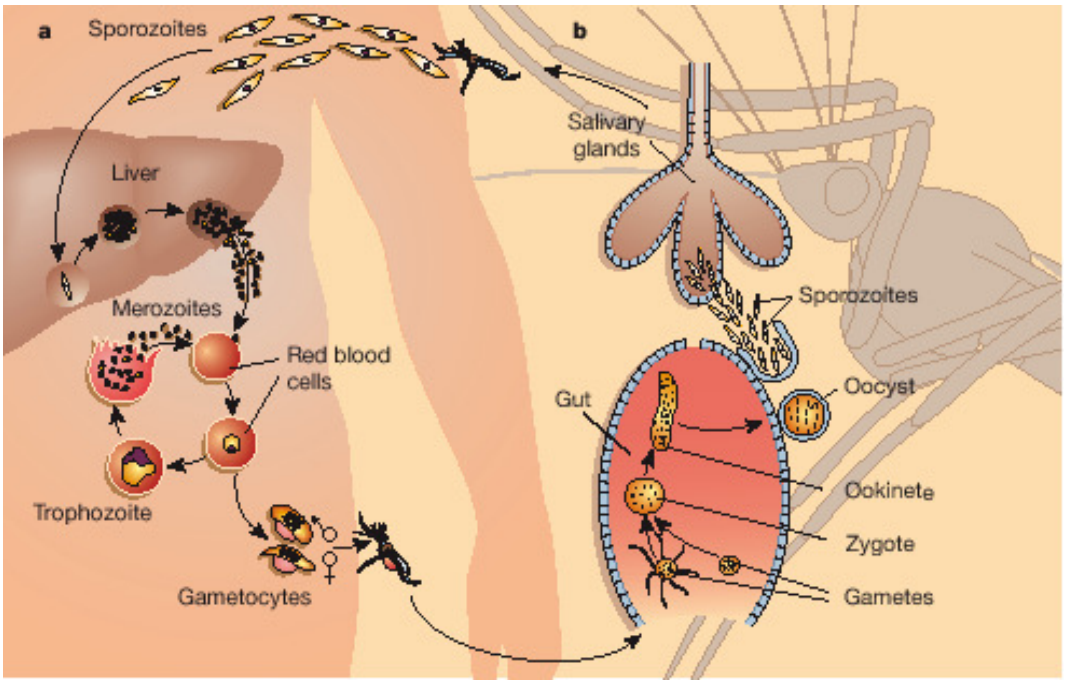


Figure 1. Life cycle of malaria parasite *P. falciparum* (Source: Wirth, 2000).

The life cycle the *Plasmodium* parasite is divided between the vertebrate host *ie* (human) where asexual replication takes place (a) and the invertebrate (Mosquito) where the sexual reproduction occurs. Details are given in the text above.

1.3 Clinical Aspects of Severe Malaria

Severe malaria is defined as an infection with manifestation and complications that are potentially fatal in man causing 15 to 20% mortality in spite of effective drugs and correct medical aid. Severe manifestation and complications due to *P. falciparum* malaria nclude a range of clinical features such as cerebral malaria, severe anemia, severe respiratory distress, hypoglycemia, renal failure and pulmonary oedema. Cerebral malaria and severe anemia are, however, the most common causes of hospitalization and death, especially in malaria naïve individuals and children (Mackintosh et al 2004). In adults, severe malaria is manifested as impaired consciousness however, multi-organ failure is more common (WHO 2000). Normally, cerebral malaria patients progressively develop coma and unconsciousness. Microvasculature occlusion by clumps of pRBCs, RBC-pRBC rosette and other fibrillar materials are believed to be the direct causes. Little is known about the exact cause of the blood-brain barrier damage.

Severe malaria in children has been considered to be primarily due to two major clinical syndromes: those with impaired consciousness (regarded as being synonymous with cerebral malaria) and those with severe malaria anaemia; of which both frequently occur in the same patient (WHO 1990). Most malaria deaths in children were previously thought to be caused by cerebral malaria and were primarily neurological in origin, or caused by severe malaria anaemia as a result of the failure to provide a blood transfusion promptly. Over the past decade, there has been increasing recognition that severe malaria is a complex syndrome affecting many organs, and that acidosis is an important component of the syndrome and the best independent predictor of a fatal outcome in both adults and children (Newton & Krishna 1998). Factors associated with fatal outcome in Kenyan children with severe malaria included deep breathing or acidosis (base excess below -8) hypotension (systolic blood pressure < 80 mmHg), raised plasma creatinine (>80 mol/l), low oxygen saturation (90 %), dehydration and hypoglycaemia (2.5 mmol/l) (Maitland et al 2003).

Acute pulmonary edema is also a common fatal complication, presenting interstitial edema with swollen endothelial cells and monocytes narrowing the capillary lumen. The edematous interstitium also contains macrophages with endocytes and malarial pigment (Duarte et al 1985). Renal failure is another important complication in severe malaria and is defined as an increase in the serum creatinine to above 3 mg/dL or an increase in blood urea above 40%. Half of the patients with renal failure present also with lung edema and 45% of these die (WHO 2000).

Laboratory data are important for diagnosis of severe malaria. Anemia (HB < 5g/dL, Ht < 20 %) is an inevitable consequence of severe malaria and jaundice (total serum bilirubin > 3mg/dL) is common to patient with renal failure and

parasitemia above 100,000/μL (WHO 2000). Another important aspect used to predict the severity of malaria is the presence of serum procalcitonin. Procalcitonin (PCT) is a known sepsis marker and is undetectable in healthy individual. A study by (Chiwakata et al 2001) concerning patients with severe malaria reported that the PCT concentration was found to be directly proportional to parasitemia. Malaria is a systemic disease whereby different systems are affected due to infection of the erythrocytes (Miller et al 2002). The signs and symptoms of severe malaria indicate a complex syndrome, established by host and parasite factors. It is believed that the virulence of *P. falciparum* is related to cytoadherence, rosetting and antigenic variations (Chen et al 2000).

1.4 Molecular aspects of Severe Malaria

1.4.1 Sequestration and Cytoadherence

During the erythrocytic stage of *P. falciparum*, mature trophozoites and schizonts sequester from the blood circulation by binding to host endothelium, predominantly in post capillary venules of the deep tissues. This phenomenon is known as cytoadherence. Dysfunction of affected organs may occur with excessive binding due to the occlusion of blood flow that causes impaired oxygen delivery. Massive sequestration in the brain is believed to be the underlying cause of coma in cerebral malaria. The reason for the parasites to sequester is unknown. However, it is speculated that they grow better in an oxygen-depleted environment than in ambient air, and binding to the endothelium is also a way to circumvent spleen-dependent destruction.

1.4.1.1 Endothelial Receptors for Adhesion

CD36 (cluster determinant 36) and thrombospodin (TSP) were the first described endothelial receptors that bound pRBCs (Barnwell et al 1985). MAbs specific to CD36 and soluble CD36 blocked the binding of pRBCs to melanoma cells and CD36-expressing COS cells (Barnwell et al 1985). By using similar approaches, *P. falciparum* receptors VCAM-1 (vascular cell adhesion molecule 1), ICAM-1 (Intracellular adhesion molecule 1), and E-selectin were later identified (Berendt et al 1989). The affinity of *P. falciparum* for binding to endothelia cell receptors is diverse, as is their role in sequestration. ICAM-1 appears to be important and has been associated to cerebral malaria (Chakravorty & Craig 2005). This receptor is present on most microvasculature surfaces and is up-regulated by TNF-α and IFN-γ, important cytokines believed to contribute to severe malaria. However, in-vitro studies have demonstrated that the affinity of most pRBCs to ICAM-1 is weak and synergetic cooperation with other receptors such as CD36 is necessary for a stable adhesion (Craig et al 1997). CD36 and TSP receptors are poorly distributed on brain endothelium. Platelet-endothelial cell adhesion molecule 1 (PECAM/CD31) had been identified as yet another endothelial receptor involved in *P. falciparum* pRBC (Treutiger et al 1997). Clinical investigations in Kenya have revealed an association between pRBC binding to PECAM receptor and severe diseases, along with reports showing that fresh isolates from children with severe *P. falciparum* malaria bind to multiple receptors (Heddini et al 2001).

The number of receptors associated with placental malaria is fewer compared to those in severe and non-severe malaria. Chondroitin sulfate A (CSA), hyaluronic acid (HA), non-immune IgG and a few unknown receptors are implicated candidates (Beeson et al 2000, Fried & Duffy 1996). CSA is the most prevalent receptor for pregnancy malaria as suggested by several studies (Duffy et al 2006, Gamain et al 2005, Tuikue Ndam et al 2005).

1.4.2 Rosetting

Rosetting is the spontaneous binding of non-infected erythrocytes to erythrocytes infected with mature asexual blood stage *Plasmodium* parasites (Rowe 2005). Its association with severe malaria has been extensively studied both clinically and experimentally by several research groups (Carlson & Wahlgren 1992, Fernandez et al 1998, Rowe et al 1997). Rosetting seems to increase microvascular obstruction of blood flow and hide the parasitized cells thereby protecting them from phagocytosis, one of the main mechanisms of anti-parasitic immunity. Previously, it was postulated that rosetting facilitates the parasite to invade uninfected RBCs (Wahlgren et al 1989). However, experiments with a culture-adapted laboratory strain PA1 showed conclusively that there was no difference in invasion rates between the isogenic rosetting and non-rosetting parasites in this strain (Clough et al 1998). All *Plasmodium spp* studied so far can form rosettes invitro (Fernandez & Wahlgren 2002, Rowe et al 2000). Different independent studies in malaria endemic areas have suggested that rosetting is associated with severe malaria (cerebral) and anemia (Fairhurst et al 2005, Newbold et al 1997, Roberts et al 2000). Nevertheless, studies in Papua New Guinea and some areas in Africa failed to correlate rosetting capability and diseases severity (Rogerson et al 1996). Host genetic factors might be the reason for the discrepancy between these studies. The rosetting capacity of parasites originating from cases of severe malaria and uncomplicated malaria were compared, parasites isolated from severe malaria displayed much higher rosetting rates than those causing uncomplicated malaria (Heddini et al 2001). In a recent study rosetting rates of fresh isolates from Ugandan children with severe and mild malaria were determined. It was clearly found that the rosetting rate of infected RBCs from severe patients was higher than that of infected RBCs of patients with mild malaria and giant rosettes were observed significantly more frequent in severe isolates (Normark et al 2007). These findings indicated that parasites causing severe disease are phenotypically

different from those causing mild disease, prompting the need to focusing studies on the composition of rosetting parasites. It was also found that serum from patients with severe malaria contained a low titer of anti-rosette specific antibodies, while patients with uncomplicated malaria had a higher titer of anti-rosette antibodies (Treutiger et al 1992). These observations indicate that anti-rosetting immunity is important factor in disease outcome.

1.4.2.1 Rosetting receptors on RBC surface

Rosetting receptors are diverse. To date, four rosetting receptors have been identified on RBCs: ABO blood group antigens, heparin sulfate-like glycosaminoglycans (HS like GAGs), CD36 and compliment receptor 1(CR1) (Chen et al 1998, Mayor et al 2005, Rowe et al 1997, Vogt et al 2003). Stable rosettes require participation of multiple serum components such as non-immune human immunoglobulins as well as other serum proteins (Luginbuhl et al 2007, Treutiger et al 1999). Oligosaccharides of the ABO blood group were the first host receptors identified in rosetting process, mainly blood group A antigens (Barragan et al 2000). Individuals with blood group A antigen phenotype are more frequently affected by severe malaria and coma than those of other blood groups. Studies have shown that rosetting is reduced in blood group O RBCs compared with the non-O groups (A, B and AB) in *P. falciparum* laboratory strains (Carlson & Wahlgren 1992) and field isolates (Udomsangpetch et al 1993). Recent evidence indicates that blood group O provides protection against severe *P. falciparum* malaria through the mechanism of reduced rosetting (Rowe et al 2007, Uneke 2007). CD36 is present in low copy numbers in mature RBCs, and it is difficult to accommodate its participation in rosetting. In contrast CR1 is widely distributed on the RBC surface, and individuals deficient in CR 1 are clinically resistance to severe malaria, as the parasites lose the capacity to form rosettes (Rowe et al 1997).

1.4.3 PfEMP 1 and *var* genes

P. falciparum erythrocyte membrane protein 1 plays a central role in host parasite interaction. PfEMP1 is expressed on the surface of infected RBCs during late stage blood infection where they mediate parasite adhesion to host cells resulting in sequestration and immuno-modulation. PfEMP1 is encoded by members of the *var* superfamily of genes that is present in about 60 copies per genome. Only one *var* gene is expressed in a single parasite at a time, switching of expression from one variant to another changes the antigenic properties and results in antigenic variation (Scherf et al 1998). Switches in *var* gene expression correlate with changes in binding pRBCs (Scherf et al 1998), which seems to allow the parasites to establish chronic infections and sequester to different sites in the body. *var* genes are found on all chromosomes except 14. The majority of *var* genes are located in subtelomeric regions and some cluster in central parts of chromosomes 4, 6, 7, 8 and 12 in 3D7 strain. *var* genes in subtelomeric regions are more vulnerable to recombination and presumably undergo frequent sequence alterations. Though the centrally located *var* genes are relatively conserved, gene recombination events affect their stability (Freitas-Junior et al 2000). As for sequence polymorphism, the sizes of *var* genes are very diverse ranging from 6-13 kb. The classical *var* genes have a two-exon structure interrupted by a 1 kb *var* intron. Exon 1 encodes an extremely diverse exposed extracellular portion which is constituted of a semi-conserved N-terminal segment (NTS) and several different domains, some of which have been shown to be responsible for binding and a predicted transmembrane (TM) domain. The second exon encodes a more conserved cytoplasmic tail (acidic terminal sequence, ATS), anchoring the protein to the knob structure on the pRBC surface (Waller et al 2002).

The extracellular binding domain of PfEMP1 is highly variable. Each PfEMP1 molecule consists of a variable number of structurally unique domains. There are four types of building blocks: the semi-conserved N-terminal segment (NTS) located on the amino terminus, the Cystein-rich InterDomain Regions (CIDR), cysteines-rich domain known as Duffy Binding-like (DBL) domains and the C2 domain (Figure 2). DBL domains are homologous to *P. falciparum* erythrocyte binding antigen (EBA) and to *P. vivax* and *P. knowlesi* Duffy binding proteins (DBP). These molecules are involved in the invasion of the RBCs by binding to RBC proteins, such as to the Duffy antigen or to glycophorin A (Gaur et al 2004). The diverse exon1 structures of PfEMP1 variants can be categorized by their domain combinations, typically ranging from 2-6 DBL sequence classes (α, β, γ, δ, ε, and x), there are two distinct types DBLα domains (α and α1) based on the number of conserved cysteines and other hydrophobic residues (Robinson et al 2003). There are only two CIDR sequence classes (α and β) and the C2 domain is conserved (Gardner et al 2002). Thirty-one different architectural types were described in the three (3D7, HB3 and IT4) sequenced parasite *var* repertoires (Kraemer et al 2007). Interestingly certain tandem domain combinations were consistently preserved e.g DBLα-CIDR1, DBLβ-c2, and DBLδ-CIDR. Most PfEMP1 variants have a semi-conserved protein head structure consisting of NTS- DBLα-CIDR1 domains (Gardner et al 2002). Sequences of DBL1α domain are relatively conserved compared to other domains within PfEMP1 but still highly diverse. DBL1α has been a target for the majority of molecular epidemiological studies of *var* gene since degenerate primers are able to amplify different sequences from laboratory clones and wild isolates from different places (Barry et al 2007, Bull et al 2005, Fowler et al 2002, Kaestli et al 2004, Kirchgatter et al 2000, Kyes et al 1997, Kyriacou et al 2006, Rottmann et al 2006, Taylor et al 2000a, Ward et al 1999). Although all *var* genes maintain a basic architecture, the amino acid sequence is highly variable when comparing PfEMP1 proteins among paralogues and across

parasite isolates. This suggests repertoire of PfEMP1 is virtually unlimited within natural parasites populations. Most PfEMP1 proteins have an overall amino acid identity less than 50% in individual domains even among proteins of the same architectural type (Kraemer et al 2007). There is minimal overlap of DBLα tags in population surveys of parasite isolates (Barry et al 2007). The vast antigenic diversity of PfEMP1 proteins in the parasite population may help to explain why individuals are repeatedly susceptible to *P. falciparum* infections and never develop sterilizing immunity. Despite the fact that the diversity of variant antigens is so large, hyperimmune human sera from distinct geographical locations are able to recognize pRBC and thus presumably PfEMP1 from East or West Africa (Aguiar et al 1992), signifying that some epitopes are semi conserved and globally related, possibly due to common ancestors.

1.4.3.1 PfEMP1 binding domains

In vitro studies have identified a range of host receptors binding pRBCs. Different domains of PfEMP1 seem to have diverse binding affinities to different endothelia receptors. The regions responsible for the binding of PfEMP1 proteins have been mapped to the DBL and CIDR domains. PfEMP1 proteins have related protein architecture, but differ broadly in sequence, domain composition and binding specificity for both protein and carbohydrate substrates (Figure 2). DBL and CIDR domains can be divided into different sequence types on the basis of sequence similarity (Smith et al 2000b). Obviously domains that bind to the same host receptor are related. For example nearly all CIDRα type domains bind to CD36, whereas CIDRβ domains do not bind to CD36 (Robinson et al 2003). DBLα binds to heparin sulphate, heparin, CR1, and blood group A (Chen et al 1998, Rowe et al 1997). DBLβc2 domain of a PfEMP1 has been mapped as an ICAM-1 binding receptor (Smith et al 2000a). In 3D7 clone, the number of PfEMP1s with potential

ICAM-1 binding domains is quite few (Rasti et al 2004). One could speculate that only a limited number of parasites can bind to this receptor. Whether these parasites only contribute to cerebral malaria remains to be elucidated. CSA binding domain has been mapped to CIDR1α, DBL3γ and DBL5δ in the PfEMP1 protein (Buffet et al 1999, Degen et al 2000, Reeder et al 1999)

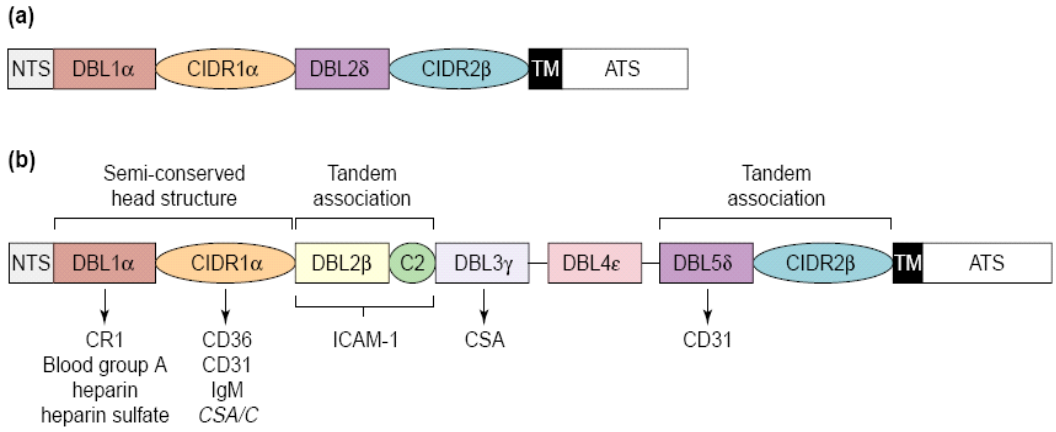


Figure 2. A schematic diagram of PfEMP1 protein architecture and binding domains

(a) The prototypical PfEMP1 extracellular regions consist of an NTS and DBLα-CIDR1 "semi conserved head structure" followed by a DBL2δ-CIDR2β tandem.

(b) Larger PfEMP1 proteins, includes the DBLβ, γ and ε types arrayed differently. Mapped binding traits for receptors are indicated with the domain that is responsible for binding (Explanation are given in the text). Adapted from (Smith et al 2001)

Based on chromosomal location, gene orientation and the conserved 5'flanking sequences in 3D7, *var* genes have been grouped into three distinct groups commonly called A, B and C (Gardner et al 2002, Voss et al 2000) with two intermediate groups (B/A and B/C) Lavstsen et al (2003). *var* group A genes are best defined, and are comprised of UpsA flanking sequences, located in sub-telomeric regions transcribed toward the telomere, encoding PfEMP1 with a complex domain structure. *var* group B is the largest group within the *var* gene family, they consists of telomeric genes flanked by UpsB sequences that are transcribed toward the centromere, and *var* group C are flanked by UpsC sequences and are located in central chromosomal regions (Figure 3). Group B/A genes are very similar in location and transcriptional orientation to group B genes, but are located further from the telomere following other *var* genes or pseudogenes. In contrast, group B/C genes have an UpsB-like 5' flanking

sequence, but are located in central chromosomal regions. Thus, it has been postulated that groups B/A and B/C represent transitional groups between the major groupings (Kraemer & Smith 2003). Both *var* group A and B/A genes are larger and have a more complex domain structure than other groups and encode a distinct protein head structure (Lavstsen et al 2003). Characteristic of *P. falciparum var* gene groups are summarized in Table 1 below

Table 1: Characteristics of *P. falciparum var* gene groups

var gene group	Upstream sequence	Position	Orientation (direction of transcription)	No. of genes in 3D7	No. of DBL domains	No. of cysteines in amplified DBL1α tag
A	UpsA	Subtelomeric	Telomeric	10	2-5	2
B/A	UpsB	Subtelomeric	Centromeric	4	4-7	2 or 4
B	UpsB	Subtelomeric	Centromeric	21	2-3	2 or 4
B/C	UpsB	Central	Telomeric	10	2-3	4
C	UpsC	Central	Telomeric	13	2-3	4

Adapted from (Kyriacou et al 2006) with modifications

Inter-isolate comparisons have revealed the existence of four unusual *var* genes: *var1csa*, *var2csa*, Type 3 *var* and *var*4 genes which appear in nearly all parasite isolates (Kraemer & Smith 2003). Type 3 v*ar* and *var*4, it became that they belong to subgroup A because they are both located in the subtelomeric region of chromosomes and are transcribed towards the telomere (Kraemer et al 2007). In field isolates *var2csa* appears to have semi-conserved homologues and it seems to play a critical role in the pathogenesis of pregnancy associated malaria. Its conserved sequence structure makes it a possible vaccine target against placenta malaria (Rowe & Kyes 2004). *var4* in 3D7 *P. falciparum* clones has been shown to be highly transcribed in severe malaria patients (Jensen et al 2004). No function has yet been ascribed to the proteins encoded by *var1csa* and Type 3 *var*.

var genes are not the only gene family localized at the teleomers. The repetitive interspersed family (*rifin*) and the subtelomeric variable open reading frame family (*stevor*) are localized adjacent to the *var* genes (Gardner et al 2002). Both families show antigenic variation and are associated with the RBC membrane, their function is not yet clear. RIFINs have been implicated in the formation of rosettes between pRBC and uninfected RBCs, however this role has yet to be verified (Kyes et al 1999).

Sequence and binding analysis of 3D7 *var* genes indicate recombinant CIDR domain based on *var* group A sequences do not bind to CD36, by contrast to CIDR domains produced on the basis of *var* group B and C (Robinson et al 2003). Thus, *var* gene recombination hierarchies may promote the evolution of PfEMP1 adhesion groups with different patterns of sequestration and disease. A fundamental question is whether the gene organization observed in 3D7 occurs in other parasite isolates and contributes to a general recombination mechanism shaping the variant antigen repertoires.

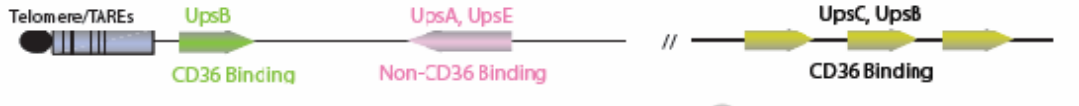


Figure 3. Chromosomal organization of *var* genes

var genes are classified according to upstream promoter type, direction of transcription and binding phenotype. Most *var* genes are found at the subtelomeric part of *P. falciparum* chromosomes and some clustered in internal regions on chromosomes. Arrows indicate the direction of transcription. TAREs: telomere associated repeat elements (Details are given in the text). Adapted from (Kyes et al 2007)

1.4.3.1 Severe Malaria and PfEMP1 expression

The clinical outcome of a malaria infection depends on multiple factors, including parasite and host polymorphisms and immune status (Mackintosh et al 2004). Severe malaria has previously been associated with expression of restricted subset of *var* genes that are antigenically conserved within the repertoires of the *var* gene family (Bull et al 2000, Nielsen et al 2002). Of the different malaria diseases syndromes, the role of PfEMP1 is best understood for pregnancy-associated malaria (PAM) Gamain et al (2007). In malaria endemic areas primigravidae women are the prime victims of placental malaria. During pregnancy, women who have previously developed malaria immunity become susceptible to infected RBCs which binds to CSA in the placenta and trigger development of PAM and other forms of severe disease. After one or two pregnancies, women develop protection to the placental form of the disease. This protection is correlated with development of antibodies that recognize placental parasite from different geographical regions, suggesting that the surface molecules expressed by placental infected RBCs may have unique and conserved features. VAR2CSA has been identified and found to be conserved across global isolates, and is transcriptionally up-regulated in placental isolates and parasites selected to bind CSA (Gamain et al 2005). Disruption of *var2CSA* causes infected RBCs to loose their ability to bind CSA.

Similar to PAM, parasites variants associated with severe childhood malaria appear to have less antigenic diversity than those associated with mild infections indicated by broader serological reactivity with semi immune children's sera (Nielsen et al 2002). The adhesive phenotypes associated with severe childhood malaria are less well defined than PAM, and the extent of PfEMP1 remains to be characterized. Severe childhood malaria encompasses several clinical syndromes (severe anaemia, cerebral malaria, respiratory distress, and hypoglycaemia and

has been linked to sequestration of pRBC to many tissues. To determine if specific PfEMP1 proteins are responsible for one or more of the severe malaria syndromes is one of the pertinent research questions within the scientific community in search for vaccine target candidate. By using different approaches researchers are analysing *var* gene expression during infections, characterizing the antibody response to the pRBC surface, and investigating the binding properties of PfEMP1 protein. Jensen and colleagues investigated the antibody reactivity of PfEMP1 protein by panning 3D7 pRBCs on semi-immune children sera. UpsA *var* transcripts were found to be up-regulated suggesting that UpsA *var* genes might contain common PfEMP1 antigenic types that are expressed in early childhood infections, this report led to proposition that UpsA genes might be associated with severe malaria and probably UpsB and UpsC *var* genes are associated with mild infections (Jensen et al 2004).

To date, only six studies have been carried out to investigate types of *var* genes sequences from field isolates that are expressed during disease, and patients with different forms severe malaria have been involved (Bull et al 2005, Kaestli et al 2006, Kirchgatter & del Portillo 2002, Kyriacou et al 2006, Montgomery et al 2007, Rottmann et al 2006). These studies are complicated by the extensive variation and simultaneous transcription of *var* genes and technical difficulties such as primer bias and clinical definition of severe malaria. Differences in epidemiology, severe disease classification, and *var* classification make comparison across studies difficult. However, with larger sample sizes and strict case definition of severe malaria it would probably give a different picture of expressed PfEMP1 sequences and disease phenotype. In three studies, expressed DBLα sequence tags were classified by the number of cysteines encoded (Kirchgatter & del Portillo 2002) as well as other features (Bull et al 2005, Kyriacou et al 2006). Genes with two cysteines in this region (2cys/DBLα1-type) are likely to represent UpsA *var* genes

or a subgroup of B/A (Kyriacou et al 2006) whereas those of four cysteines are either UpsB or UpsC. The expression of (2cys/DBLα1-UpsA) sequence variant correlates with rosetting phenotype (Bull et al 2005), with cerebral malaria in Malian children (Kyriacou et al 2006) and non cerebral severe malaria in adults (Kirchgatter & del Portillo 2002). A study in French Guyana involving 19 severe malaria patients demonstrated that there were particular DBLδ *var* sequences expressed exclusively by these patients as opposed to 32 patients with mild malaria (Ariey et al 2001). Similar results were obtained in Brazil where parasites from patients with severe malaria transcribed predominantly DBL1α *var* sequences lacking 1-2 cysteines residues, while parasites from patients with mild malaria transcribed preferentially DBL1α *var* sequences without these deletions (Kirchgatter & del Portillo 2002). In our previous case control study in Ifakara, Tanzania, using quantitative Real-time PCR analysis we reported a correlation between expression of both UpsA and UpsB *var* expression and severe malaria cases in children (Rottmann et al 2006). In Papua New Guinea (PNG), our group reported only UpsB *var* expression correlated with severe disease (Kaestli et al 2006). However, in PNG, 80% of the population is deficient in CR1, a major receptor for pRBC rosetting. Additionally in PNG, rosette phenotype does not relate with severe disease. Human genetic polymorphisms in cytoadhesion receptors may influence PfEMP1 disease associations. Strict correlations between any group of *var* genes and disease manifestation have not been found.

Different parasite genotypes are potentially virulent, severe malaria syndromes are relatively infrequently complication of malaria infections suggesting that isolate-transcendent disease immunity can develop rapidly. PfEMP1 immunity is an important factor in the rapid development of disease immunity. The variant antigen within the *var* genes family is vast, serological evidence suggests that the variant antigens associated with disease may be antigenically restricted (Bull et al 2000).

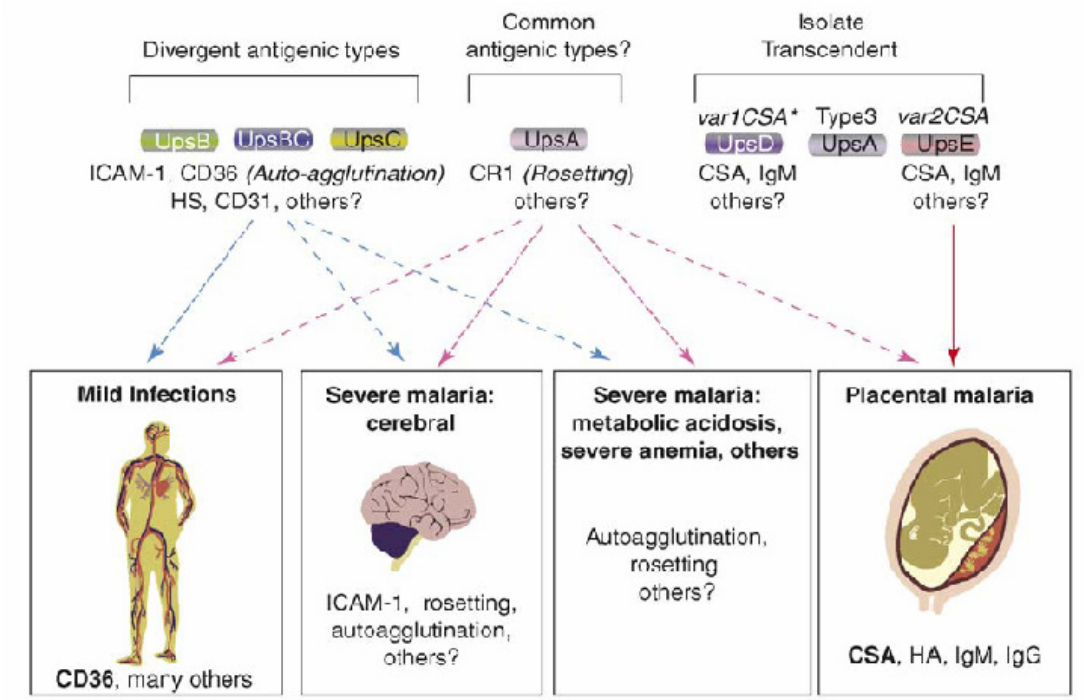


Figure 4. PfEMP1 expression and disease phenotype

This schematic figure summarizes the PfEMP1 grouping and their potential role in infection and diseases (detailed explanation are given in the text above). Dashed and solid lines indicate hypothetical infection outcomes; line density represents the level of confidence in the prediction. Adapted from (Kraemer & Smith 2006).

However, this current understanding of cytoadherence and morbidity association has become even more complex with new finding of Montgomery and colleagues, who tested PfEMP1 expression in postmortem organs and tissues. Their most intriguing finding was that parasites of one child expressed several *var* genes in a tissue specific manner (Montgomery et al 2007).

1.5 Antigenic variation of *var* genes

Survival of the malaria parasite in human host is constantly challenged by host the immune system. *P. falciparum* has developed a process called clonal antigenic variation which allows the parasite to overcome immune attacks by periodically changing the antigenic phenotype at the surface of pRBCs as reviewed in (Kyes et al 2001). Switching of surface expression can alter the interaction with host tissues which is highly linked to pathogenicity. *var* genes have been shown to be expressed in a mutually exclusive manner at both the mRNA and protein level (Voss et al 2006). By limiting expression to a single *var* gene copy the parasite limits exposure to a single antigen at a time within the host's immune system. Over time the immune system generates an antibody response against the surface of the pRBC, thus recognizing the predominantly expressed form of PfEMP1 and consequently selecting for subpopulations of parasites that arise via switching expression to different *var* genes. In malaria infections, especially in chronic cases, the peaks of parasitemia may fluctuate over time, apparently this phenomenon is inexhaustible in single infection, and is normally characterized by oscillations of peripheral parasitemia having magnitudes varying from undetectable levels to high parasite burden (Miller et al 1994). Recrudescence with the appearance of different antigenic parasites is an essential strategy for malaria parasite survival. Switching expression into different *var* genes makes the host immune response frequently inefficient, leading to prolonged infection. Malaria parasites challenge the immune system in at least two ways; through genetic recombination in the mosquitoes, a process which results in unlimited changes of the malaria genomic pool in the wild, and through the existence of several variable antigenic families.

Chapter Two

General Objectives and Study Population

Chapter Two

2.0 General Objectives and Study Population

2.1 Study Goal

Clinical manifestations of *falciparum* malaria differ remarkably from infection to infection; disease symptoms often remain mild, in most cases only a small proportion of patients develop severe complications, which encompasses a variety of clinical syndromes such as cerebral malaria, severe malaria anaemia, metabolic acidosis, respiratory distress or prostration. These syndromes may have different underlying pathogenic mechanisms. The reasons why certain children develop life-threatening complications, whilst the majorities are able to tolerate high parasite densities without severe clinical symptoms remain elusive. Severe malaria has previously been associated with expression of a restricted and antigenically conserved subset of variant erythrocyte surface antigens (Bull et al 1999), suggesting that expression of certain *var* genes may be associated with specific disease manifestations. Several sero-epidemiological studies have shown that antigens associated with severe disease are frequently recognized by sera from semi immune individuals (Bull et al 2000, Nielsen et al 2002). This suggests that expression of a particular surface molecule may be associated with specific disease manifestations.

The main goal of this thesis was to explore the relationships between sequence and disease phenotype by analysing and comparing expressed *var* gene fragments of different *Plasmodium falciparum* isolated from Tanzanian children with different manifestations of malaria (severe and asymptomatic malaria) in order to determine whether there are conserved *var* sequences motifs that are associated with the severity of the disease.

2.2 Specific Objectives

1. To identify expressed *var* genes in naturally infected individuals
2. To identify expressed *var* genes associated with severe malaria through a case control study
3. To determine the *var* gene distribution in clinical and asymptomatic cases
4. To determine the phylogenic relationship of sequences generated from severe and asymptomatic isolates
5. To estimate the repertoire of expressed *var* genes

2.3 Study area and population

2.3.1 Study area

These studies were carried out in Ifakara, a semi rural area in south – eastern Tanzania from June to September 2003 and January 2005. Ifakara is an area of moderate perennial *P.falciparum* transmission surrounded by areas of intense transmission. Patients with severe malaria were recruited in a pediatric ward of Saint Francis Designated District Hospital (StFDDH). Children with asymptomatic malaria were recruited from a retrospective survey in nearby villages surrounding the hospital. Data from the hospital showed that 40 % of the children admissions are due to infection with *P.falciparum*. Malaria is reported to account for a case fatality rate of 2.4 % in this hospital (Schellenberg et al 2004).

2.3.2 Study Population and Recruitment Criteria

Ifakara Health Research & Development Centre (IHRDC) is closely linked to StFDDH. This link provides the basis for a comprehensive clinical surveillance system and allows the research activities on severe childhood illnesses. Children less than 5 years of age presenting severe malaria at the hospital were recruited in the study after obtaining the informed consent from the children's parents or guardians. Severe malaria cases were defined according to the world health organization (WHO) criteria for severe malaria (WHO 2000). In brief, a child is considered to have severe malaria if had a *P. falciparum* positive smear as the primary diagnosis and no other cause for illness and in addition had one of the following symptoms: (1) prostration (2) respiratory distress or (3) severe anaemia. In this study recruitment criteria were restricted to patients admitted with cerebral malaria (unrousable coma with a Blantyre score ≤ 3 and no other detectable cause of coma). Exclusion criteria were confirmed co-infections, malnutrition (mid-upper arm circumference [MUAC] of ≤ 12 cm), haemoglobin ≤ 5 g/dL, lactate ≥ 5 mmol/L, glucose ≤ 2.2 mmol/L) or antimalarial treatment during the past 14 days. Clinical and epidemiological data were collected using inpatient forms (see appendix 1). The control group consisted of children with asymptomatic malaria (presence of *P. falciparum*, axillary temperature of ≤ 37.5 °C and no other symptoms), and were recruited by a convenience sampling in the nearby villages after obtained informed consent from the children's parents or guardians. These children were the screened for the presence of malaria parasites by using rapid diagnostic test. Children with *P. falciparum* infection were then confirmed at IHRDC laboratory by Giemsa-stained blood films.

2.4 Ethical Considerations

The proposal for this study was reviewed and approved by Ifakara Health Research and Development Centre Institution Review Board (IRB) and the Medical Research Coordinating Committee of the National Institute for Medical Research in Tanzania.

Chapter Three

Differential Expression of *var* Gene Groups Is Associated with Morbidity Caused by *Plasmodium falciparum* Infection in Tanzanian Children

INFECTION AND IMMUNITY, July 2006, p. 3904–3911 Vol. 74, No. 7
0019-9567/06/$08.00+0 doi:10.1128/IAI.02073-05

Differential Expression of *var* Gene Groups Is Associated with Morbidity Caused by *Plasmodium falciparum* Infection in Tanzanian Children

Matthias Rottmann,[1]† Thomas Lavstsen,[2]† Joseph Paschal Mugasa,[3] Mirjam Kaestli,[1] Anja T. R. Jensen,[2] Dania Müller,[1] Thor Theander,[2] and Hans-Peter Beck[1]*

Swiss Tropical Institute, Socinstrasse 57, Postfach 4002 Basel, Switzerland[1]; Centre for Medical Parasitology at Institute for Medical Microbiology and Immunology, University of Copenhagen, Panum Institute 24-2, Blegdamsvej 3, 2200 Copenhagen N, Denmark[2]; and Ifakara Health Research and Development Centre, P.O. Box 53, Ifakara, Morogoro, Tanzania[3]

Received 23 December 2005/Returned for modification 28 January 2006/Accepted 4 April 2006

The *var* gene family of *Plasmodium falciparum* encodes the variant surface antigen *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1). PfEMP1 is considered an important pathogenicity factor in *P. falciparum* infection because it mediates cytoadherence to host cell endothelial receptors. *var* genes can be grouped into three major groups, A, B, and C, and the conserved *var* genes, *var1-4*, according to sequence similarities in coding and noncoding upstream regions. Using real-time quantitative PCR in a study conducted in Tanzania, the *var* transcript abundances of the different *var* gene groups were compared among patients with severe, uncomplicated, and asymptomatic malaria. Transcripts of *var* group A and B genes were more abundant in patients with severe malaria than in patients with uncomplicated malaria. In general, the transcript abundances of *var* group A and B genes were higher for children with clinical malaria than for children with asymptomatic infections. The *var* group C and *var1*-like transcript abundances were similar between the three sample groups. A transcript abundance pattern similar to that for *var* group A was observed for *var2csa* and *var3*-like genes. These results suggest that substantial and systematic differences in *var* gene expression exist between different clinical presentations.

The particular virulence of *Plasmodium falciparum* is linked to the cytoadhesion properties of infected erythrocytes in deep vascular beds leading to multiple complications and symptoms (19). This process of sequestration is thought to be an immune evasion strategy to avoid splenic clearance (2, 8, 14). Infected erythrocytes also form rosettes with uninfected erythrocytes (27) or form larger groups involving platelets, called clumps (25). *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), encoded by the *var* gene family and expressed on the surfaces of infected erythrocytes, mediates binding to host endothelial receptors and is an important target for protective immunity (1, 7, 9, 13, 31, 34). Each parasite possesses 50 to 60 *var* gene copies, and switching between surface expression of the various *var* gene products results in antigenic variation while maintaining or changing adhesion properties (12, 31). Immunity preventing severe malaria and death develops naturally in exposed populations. In areas of intense transmission, the main burden of malaria morbidity and mortality is among children between 6 months and 5 years of age (32). Adults are often infected asymptomatically, and severe disease is rare.

It has been shown that parasites from patients with severe malaria express a different subset of surface antigens that are more frequently recognized by sera from malaria-exposed individuals, including young children, than parasite antigens from older children with mild malaria (5, 23). It has also been shown that this subset of surface antigens is serologically conserved among different geographical regions (24), and it is therefore crucial to identify the molecular phenotype of such a subset to develop a disease-ameliorating vaccine.

Linking PfEMP1 expression to disease outcome is inherently difficult due to the extensive inter- and intragenomic variation in *var* genes. Previous studies predominantly relied on reverse transcription-PCR using degenerate primers, with subsequent cloning and sequencing (4, 17). While these studies have shown that the transcription of certain DBL1α domains is associated with either severe malaria or rosette formation, they have been unable to identify a clear correlation between *var* gene groups and disease outcomes.

Sequencing of the 3D7 genome revealed that *P. falciparum* parasites contain 50 to 60 *var* genes that can be grouped into three major groups, A, B, and C, and the single-copy intergenomic conserved *var* genes, *var1* and *var2csa*, according to sequence similarities in both noncoding and coding sequences (12, 18, 22, 37). Evidence is emerging for the existence of subgroups of *var* group A, namely, type 3 *var* (18) and type 4 *var* (12, 15) genes, referred to here as *var3* and *var4*, respectively. The functional relevance of this genetic structuring is indicated by the fact that CIDR domains of group B and C PfEMP1 variants bind to CD36, in contrast to CIDR domains of group A variants (26). Parasites selected for chondroitin sulfate A and human bone marrow endothelial cell binding in vitro dominantly express *var2csa* and *var4*, respectively (15, 29), also supporting the notion of functional genetic substructuring.

The genetic organization of *var* genes was exploited to de-

* Corresponding author. Mailing address: Swiss Tropical Institute, Socinstrasse 57, CH 4002 Basel, Switzerland. Phone: 41-61-284 8116. Fax: 41-61-284 8101. E-mail: hans-peter.beck@unibas.ch.

† M.R. and T.L. contributed equally to this work.

TABLE 1. SYBR green primers

Primer pair	Sequence (5′–3′)	
	Forward	Reverse
A1	TTGGGRAATBTGTTAGTTAYRGCAA	CTGCAAAACTKCGWGCAAG
A2	AACCCATCTGTRRATGATATACCTATGGA	GTTCCAASGATCCATTRGATGTATTA
A3	AGGTAATGTTTTAGATGATGGTAT	ACCAGAATATACATTATTTGATACATA
B1	CATCCGCCATGCAAGTATAA	CGTGCACGATTTCGATTTTT
B2	ATCAAGGTAATTTCATACATATGTGATA	GTCCGTGCACGATTTCGATTTT
C1	CACATCGATTACATTTTAGCGTTT	TGTGGTAATATCATGTAATGG
C2	GTAGCGACAACCACGRYATCATGG	CATTGTTAACATAGTCTACCATTA
BC1	GACAAAACTTTCACCCAATAGA	AATGATCGGTGTAACCACTATC
BC2	CATCTGTTGCAAATTTATTCCAAATAC	TCAGTAGTATCAGACATAAATGCATA
pvar1utr	TGGCACATCTTTGGTATAAAA	AAACCTTTATATTCCTGTAAAATTCA
pvar2utr	CACGACATTAACAATACATGCAGA	CATTGCATTCACAGACATTGG
pvar3coding	CGTAAAACATGGTGGGATGA	GGCCCATTCAGTTAACCATC

sign primers targeting the conserved regions defining *var1-3* genes and group A, B, and C *var* genes. Using these primers in quantitative reverse transcription-PCR, the transcript abundances of *var* genes were measured in parasites collected from *P. falciparum*-infected Tanzanian children with asymptomatic malaria (AM), uncomplicated malaria (UM), and severe malaria (SM). Our data demonstrate an increase in transcript abundance for group A and B *var* genes in parasites causing severe malaria compared to that in parasites causing uncomplicated malaria.

MATERIALS AND METHODS

Study design and population. The study was conducted in Ifakara, a semirural area in southern Tanzania, from June to September 2003. Ifakara is an area of moderate perennial *P. falciparum* transmission surrounded by areas of more intense transmission.

Of all children seeking medical treatment, 40% were admitted to the hospital due to infection with *P. falciparum*. Malaria is reported to account for a case fatality rate of 2.4% in this hospital (30). Samples were collected from children (aged 4 to 59 months) presenting with malaria at the hospital. Severe malaria cases were defined according to the World Health Organization criteria for severe malaria (38). Uncomplicated malaria was defined as the presence of asexual *P. falciparum*, an axillary temperature of >37.5°C, or symptoms of headache or myalgia but no other signs of severe malaria. Exclusion criteria were confirmed coinfection, malnutrition (mid-upper-arm circumference [MUAC] of <12 cm), or antimalarial treatment during the last 14 days. Asymptomatic patients (presence of *P. falciparum*, axillary temperature of <37.5°C, and no other symptoms) were age-matched as closely as possible to the patients with severe cases by convenience sampling in the same area in January 2005. *P. falciparum* infection was determined by Giemsa-stained blood films, and parasitemia was counted as parasites per 200 white blood cells. Ethical clearance for this study was obtained from the Ifakara Health Research and Development Centre's scientific review board and the Medical Research Coordinating Committee of the National Institute for Medical Research in Tanzania.

Blood samples. After obtaining written informed consent from parents, 1 to 2 ml of venous blood from children was collected in EDTA tubes. Erythrocytes were separated from serum by centrifugation and washed with 40 ml phosphate-buffered saline, 5 volumes of TRIzol reagent (Invitrogen) was added to the erythrocyte pellet, and the sample was frozen at −70°C. Samples were collected from 52 patients with SM, 56 patients with UM, and 19 AM children. 3D7 parasite lines were generated as described elsewhere (33).

DNA, RNA, and cDNA. Genomic DNAs were isolated from infected red blood cells (100 μl blood was frozen after adding 2 volumes of 6 M guanidine HCl, 50 mM Tris, pH 8.0, 20 mM EDTA) with a QIAamp blood kit (QIAGEN), and total RNAs were extracted by using TRIzol reagent (Invitrogen) twice, as recommended by the manufacturer, and treated with DNase I (Invitrogen) for 30 min at 37°C. The absence of DNA in RNA samples was confirmed by stable base fluorescence after 40 cycles of real-time PCR with seryl-tRNA synthetase primers (29). Reverse transcription was performed using Superscript II (Invitrogen) and random hexamer primers in a total volume of 40 μl according to the manufacturer's recommendations. Hereafter, two different real-time PCR methods were used to quantify *var* transcript abundance, namely, a SYBR green-based assay and a minor groove binder (MGB) probe-based assay.

Validation of SYBR green method and quantification of *var* gene transcript abundance. Quantitative real-time PCR using Quantitect SYBR green PCR master mix (QIAGEN) was performed on a Rotorgene thermal cycler system (Corbett Research) as previously described, using the seryl-tRNA synthetase (primer pair p90) and fructose-bisphosphate aldolase (primer pair p61) genes as endogenous controls (23). Primers targeting *var* gene groups A, B, and C were designed based on sequence similarities in the 3D7 *var* repertoire (22). The majority of *var* genes are flanked by conserved upstream region upsA, upsB, or upsC. Primer pairs B1 and B2 target the conserved upstream region of *var* B genes, whereas C1 and C2 target the upstream region of *var* C genes. Attempts to design primer pairs targeting upsA with sufficient amplification in the SYBR green assay were unsuccessful. In strain 3D7, grouping into groups A, B, and C is also maintained in the coding sequences for DBLα and ATS. This was exploited to design primer pairs A1, targeting DBLα of group A *var* genes; A2 and A3, targeting exon 2 of group A *var* genes; and BC1 and BC2, targeting exon 2 of both B and C *var* genes. Additional primers were designed to target conserved regions of *var1* and *var2csa* 5′ untranslated regions (UTRs) and *var3* coding regions (pvar1utr, pvar2utr, and pvar3). Primers designed to target *var4* genes in quantitative reverse transcriptase PCR amplified fragments with the expected melting temperature (T_m) for 3/20 genomic DNAs from field samples only, and this analysis was therefore left out subsequently. Seven annotated genes in 3D7 were predicted not to be targeted by any of these primer pairs. Primers are shown in Table 1, and the 3D7 genes expected to be amplified by the respective primers are listed in Table 2. The primers were validated as follows. Initially, all primers were tested on 10-fold dilutions of 3D7 genomic DNA (gDNA). All primers amplified fragments of the expected size and T_m, and sequencing of one PCR clone from each amplification reaction revealed an expected target sequence. All primers had amplification efficiencies (E) between 1.85 and 2 [$E = 10^{(-1/\text{slope of 10-fold-dilution gDNA standard curve})}$] (data not shown). Cycle threshold (C_T) values for primers targeting multiple genes were compared to those obtained with the control primer p90, targeting a single-copy gene. The observed ΔC_{Ti}($\Delta C_{Ti\text{-obs-3D7}} = C_{Ti} - C_{Tp90}$) was compared to the expected value estimated from the number of predicted genes targeted in 3D7. A ΔC_T value reduction of 1 represents a duplication of targeted gene copies, and the estimated $\Delta C_{Ti\text{-estl}} = -\log(\text{no. of targeted genes})/\log(2)$ (Table 2). Most primer pairs amplified as expected; the primers were then tested on gDNAs from 20 field isolates, and the amplification results were compared to the p90 C_T values ($\Delta C_{Ti\text{-obs-field}} = \mu_{n-20}\Delta C_{Ti} - C_{Tp90}$) (Table 2). All primer pairs yielded fragments with the sizes and T_ms expected from the amplifications of 3D7 gDNA, and all primers amplified with similar efficiencies (95% confidence interval for $\mu_{(n=20)[C_{Ti}-C_{Tp90}]}$, <0.8) from all field isolates, with no significant difference (t test) between DNAs from severe and uncomplicated malaria cases. The amplification efficiency for field isolate gDNA was similar to that for 3D7 gDNA for most primer pairs, except C1, C2, and BC2, indicating that these primers targeted fewer genes in field isolates than in 3D7 and might reflect that a larger variation in *var* group C gene copy numbers exists among parasites from naturally infected individuals. However, none of the primer pairs showed any significant difference in amplification efficiencies between DNAs from severe and uncomplicated malaria cases.

Next, primers targeting *var* groups were tested on cDNAs from isogenic but phenotypically distinct 3D7 parasites with known differential *var* transcript abun-

TABLE 2. Technical characteristics of *var* group primers for SYBR green real-time PCR[a]

Characteristic	Value or description for indicated primer pair				
	A1	A2	A3	B1	B2
Target region	Grp A DBL1α	Group A exon 2	Group A exon 2	upsB	upsB
Targeted genes in 3D7 predicted by alignments	PFD1235w, MAL7P1.1, PF08_0141, PF11_0008, PFD0020c, PF11_0521, PF13_0003, PFA0015c, MAL6P1.314, PFI1820w	PFD1235w, MAL7P1.1, PF08_0141, PF11_0008, PFD0020c, PF11_0521, PF13_0003, PFA0015c, MAL6P1.314, PFI1820w	PF11_0521, PFD1235w, MAL7P1.1, PF13_0003, PFD0020c, PF11_0008	MAL6P1.1, PF07_0139, PF08_0142, PF10_0001, PF10_0406, PF11_0007, PF13_0001, PF13_0364, PFA0005w, PFA0765c, PFB0010w, PFB1055c, PFC0005w, PFC1120c, PFD0005w, PFD1245c, PFE0005w, PFI0005w, PFI1830c, PFL0005w, PFL0935c, PFL2665c	MAL6P1.1, PF07_0139, PF08_0142, PF10_0001, PF10_0406, PF11_0007, PF13_0001, PF13_0364, PFA0005w, PFA0765c, PFB0010w, PFB1055c, PFC0005w, PFC1120c, PFD0005w, PFD1245c, PFE0005w, PFI0005w, PFI1830c, PFL0005w, PFL0935c, PFL2665c
Coverage in 3D7 (group)	All A genes, including *var3* types	All A genes, including *var3* types	All A genes, excluding *var3* types	All B genes	All B genes
Fragment T_m (°C) in 3D7	77.2	73.6	72.1	75.4	75.6
Fragment length (bp)	110–120	100	160	260	190
Fragment T_m (°C)[b] in field isolates (SD) [no. of PCR-negative genomes/total]	77.2 (0.4) [0/20]	73.7 (0.2) [0/20]	72.2 (0.4) [0/20]	75.1 (0.4) [0/20]	75.3 (0.4) [0/20]
$\Delta C_{T_{\text{i-estimated-3D7 gDNA}}}$ [−log(no. of targeted genes)/log(2)]	−3.3	−3.3	−2.6	−4.5	−4.5
$\Delta C_{T_{\text{observed-3D7 gDNA}}}(C_{T_i} - C_{T_{p90}})$	−2.0	−2.5	−1.9	−3.7	−3.9
$\Delta C_{T_{\text{i-mean-observed-field gDNA}}}$ [$\mu(C_{T_i} - C_{T_{p90}})$ (95% CI)][c]	−1.2 (0.4)	−1.9 (0.3)	−1.9 (0.3)	−4 (0.2)	−3.6 (0.2)

[a] The following 3D7 *var* genes were not targeted by any primer pair: PFD0635c, PF07_0050, PFL1955w, MAL7P1.55, MAL6P1.4, MAL6P1.316, and PFL0020w (three BA and four BC genes).

[b] Mean *Tm* based on real-time PCR on 20 gDNA and 108 cDNA field isolates.

[c] 95% confidence intervals based on real-time PCR on gDNAs from 20 field isolates. The design and validation of primers and probes for the MGB real-time PCR assay were described by Kaestli et al. (16).

dance patterns. These were nonmanipulated $3D7_{UM}$ and $3D7_{SM}$ selected on hyperimmune serum (15). The transcription measured by group-specific primers was compared to predicted changes calculated from absolute quantifications using gene-specific primers. There was a clear association ($R = 0.904$; $P = 0.0008$ [Pearson correlation]) between results obtained with gene-specific and group-specific primers (Fig. 1).

Assessment of T_ms of fragments generated from cDNAs from collected field samples showed that all primers amplified fragments with the expected T_ms in 90 to 99% of all included samples (not shown), indicating that the primers targeted *var* sequences conserved in the parasite isolates. The C_T values for the two internal control genes showed with cDNAs from field samples that reliable quantification could be performed from the collected samples [$\Delta C_{T_{p90}}(\mu, \sigma) = 21.2$ and 2.6; $\Delta C_{T_{p61}}(\mu, \sigma) = 19.0$ and 2.5]. As expected, there was a negative association between parasitemia and control gene p90 C_T values (representing the overall amount of cDNA) (regression coefficient$_{[\log2(\text{parasitemia/blood sample volume}): C_{T_{p90}}]}$, −0.58; $P = 0.005$).

Validation of MGB method and quantification of *var* transcript abundance. Quantitative real-time PCR using MGB probes was performed using an ABI PRISM 7200 sequence detection system (Applied Biosystems) as described by Kaestli et al. (16), with few modifications. Briefly, cDNAs were synthesized from total RNA, and a primary PCR with 16 cycles over the *var* 5′ UTR-DBL1α target sequence was performed prior to real-time PCR for *var* groups A, B, and C. Primers and probes targeting upsA, -B, and -C for the MGB probe assay were described by Kaestli et al. (16) (upsA-probe, upsB-probe, and upsC-probe). The seryl-tRNA synthetase internal control gene was used for relative quantification without prior amplification. The primers and probe were designed using Primer Express software 2.0 (Applied Biosystems) and had the following sequences: primer p90Probe_for, 5′-ACCTCAGAACAACCATTATGTGCTT-3′; primer p90Probe_rev, 5′-TGTGCCCCTGCTTCTTTTCTAA-3′; and p90Probe, 5′-6-carboxyfluorescein-AGGTTACCACTCAAATACGCTGGATTCTCATCTTG-6-carboxytetramethylrhodamine–6-carboxyfluorescein-3′.

Data analysis. After all samples had been subjected to real-time PCR, the data set was cleaned for subsequent statistical analysis. Data points were not considered if the T_m diverged more than 1°C from the expected value or if the C_T value was above 30. Transcript abundances were compared between clinical groups after normalization to internal controls (yielding ΔC_T values) (Fig. 2). Based on these, *x*-fold changes were calculated by the $\Delta\Delta C_T$ method (see Table 4). Comparisons between groups were made with one-way analysis of variance and Intercooled Stata 8.0 analysis software.

RESULTS

Sample collection and clinical data. Samples were collected from 52 children admitted to hospital with SM and 56 children with UM. *var* gene transcription analysis was performed on cDNAs from 42 SM and 52 UM cases. Twelve samples from 19 children with asymptomatic *P. falciparum* infections collected during a village survey could be analyzed for *var* transcript abundance. Clinical characteristics of all children from whom cDNAs were available are presented in Table 3.

Comparison of *var* transcript abundance profiles. The *var* group A transcript abundance was lowest in AM cases, higher in children with UM, and highest in children suffering from SM (Fig. 2; Table 4). Similar findings were obtained with primers targeting the upsB upstream region, whereas data obtained with primers targeting group C *var* genes indicated that group C *var* genes were transcribed at the same level in the three groups of children.

The BC1 primer pair was predicted to predominantly amplify fragments of B *var* genes, whereas BC2 primers were expected to amplify a smaller subset of genes consisting of group B, BC, and C *var* genes (Table 2). The BC1 primer results showed that the targeted genes were transcribed at higher levels by SM parasites than by UM or AM parasites. No significant changes were observed in transcription of genes targeted by the BC2 primer pair.

TABLE 2—*Continued*

Value or description for indicated primer pair						
C1	C2	BC1	BC2	pvar1utr	pvar2utr	pvar3
upsC subtype PFD0625c, PFL1960w, PF07_0048, PFD0630c, PF07_0049, PF07_0051, PFD0615c, PFD1015c	upsC subtype PFD0615c, PF07_0051, PFD0995c, PFD1000c, PF07_0049, PFD1015c, MAL6P1.252	Group B and C exon 2 PF08_0142, MAL6P1.1, PFA0765c, PFC1120c, PFC0005w, PFE0005w, PF10_0406, PFB1055c, MAL7P1.56, PF08_0140, PF13_0364, PF13_0001, PFL2665c, PFI0005w, PFB0010w, PF11_0007, PF07_0139, PFD1005c, PFD1000c, PF08_0107, PFD0005w, PFL0005w, PFD1015c, PFI1830c, PF07_0050, PF10_0001	Group B and C exon 2 MAL7P1.50, PF07_0048, PF08_0103, PF08_0106, PFA0005w, PFD1245c, PFL0935c, PFL1950w	*var1* 5′ UTR PFE1640w	*var2* 5′ UTR PFL0030c	*var3* coding PFA0015c, MAL6P1.314, PFI1820w
8/13 C	7/13 C	17/22 B, 4/13 C, 1/4 BA, 2/9 BC	4/9 BC, 3/22 B, 1/13 C	*var1*	*var2*	*var3*
74.1	73.5	75.9	75.5	69.6	74	75.5
106	120	110	170	87	184	155
74.2 (0.4) [0/20]	73.5 (0.3) [0/20]	75.9 (0.3) [0/20]	75.6 (0.3) [0/20]	69.6 (0.2) [0/20]	74.3 (0.4) [0/20]	75.6 (0.2) [0/20]
−3	−2.8	−4.7	−3	0	0	−1.6
−2.2 0.3 (0.8)	−0.7 0.9 (0.5)	−4.5 −4.55 (0.2)	−1.67 0.7 (0.3)	2.8 2.6 (0.2)	−0.8 0.3 (0.3)	−1.8 ND

The primer pairs targeting the *var1* and *var2csa* (Table 1) gene family showed no significant difference in transcript abundance between the cohorts, although a trend of increased *var2csa* transcript abundance with increased severity of disease was observed. In contrast, primers targeting the *var3* family showed significantly higher transcript abundance in SM than in AM samples and a trend of higher transcript abundance with increased severity of disease (Fig. 2). *var3* belongs to the group A *var* genes, and a correlation between the transcript abundances of *var3* and group A *var* genes would be expected. The strongest correlation of *var3* transcript abundance was found with transcripts measured by A2 ($R_{\text{pvar3:A2}} = 0.482$; $P = 0.0012$ [Spearman rank]).

High transcript abundances of group A and B *var* genes are associated with severe disease. The association between transcript abundance and clinical presentation of malaria was tested in logistic regression models in which the dependent variable was the clinical presentation (SM or UM) and the independent variables were age, MUAC, parasitemia, sex, and transcript abundance measured by the respective primer pair. The logistic regression models were built for primer sets exhibiting statistically significant differences in transcript abundance (Table 4) and showed that young age and increased MUAC significantly increased the risk of severe disease. No significant association was found for parasitemia and sex (data not shown). According to the model, the risk of severe malaria is increased 20 to 61% with a twofold increased *var* group A or *var* group B transcript abundance (Table 5).

Seven severe cases were classified as cerebral malaria due to Blantyre scores of ≤3 (data not shown). Parasites from these children showed a trend towards a larger abundance of *var* group A transcripts than those for all other clinical cases (for upsA-probe, $P = 0.1660$; for A2 primer, $P = 0.0077$). Primer pairs A1 and A3 showed no difference in transcript abundance between the groups (data not shown).

Associations of *var* transcript abundance with other clinical features. Linear regression models showed that increased *var* group B transcript abundances measured by both primer pairs B1 and B2 were positively associated with parasitemia. This was also the case when corrected for age [$R_{(\text{parasites/200 leukocytes}):\Delta C_T(\text{B1/B2})} = -226/-223$; $P = 0.045/0.019$; an increase of parasitemia by 226 parasites/200 leukocytes resulted in a twofold increase in *var* group B transcript abundance (1 C_T value decrease)]. In contrast, there was a nonsignificant trend for *var* group C transcript abundance, as measured by C1 or C2 primers, to be negatively correlated with parasitemia [$R_{(\text{parasites/200 leukocytes}):\Delta C_T(\text{C1/C2})} = 224/207$; $P = 0.091/0.074$). There was also no association between *var* group A transcript abundance and parasitemia, but *var* group A transcript abundance tended to decrease with age among the UM cases ($R_{\text{age (months)}:\Delta C_T(\text{A1/A2/A3})} = 1.58/1.67/1.72$; $P = 0.050/0.058/0.081$). No such trends were found with *var* group B or C.

DISCUSSION

Studies on the development of natural acquired immunity have suggested a genetic structuring of the PfEMP1 protein family leading to niche characteristics with regards to host receptor binding (6, 7, 9, 13). This is supported by *var* expression analyses of in vitro manipulated parasites (15, 31). The present study aimed to analyze the expression of *var* genes in naturally infected individuals presenting with different forms of malaria. Based on the sequence of the *P. falciparum* clone 3D7, it appears that most *var* genes fall into one of three main groups, A, B, and C, according to both coding and noncoding regions. Four interclonally conserved *var* genes (*var1-4*) have also been identified. This genetic structuring might reflect

functional differences of the encoded PfEMP1 proteins, a notion that is supported by observations with regards to CD36 binding properties of CIDR domains (26), differences in survival rates in vivo (21), and variant surface antigen serotypes of genotypically identical parasites with diverse PfEMP1 expression profiles (15, 29). However, only a few studies have aimed to directly correlate *var*/PfEMP1 expression in naturally infected individuals with different presentations of malaria. One study showed that parasites from Chinese children suffering from cerebral malaria expressed larger PfEMP1s than did parasites from other malaria patients (3). This may indicate an involvement of group A genes in severe malaria, as large size is characteristic of but not unique to group A PfEMP1s (22).

PfEMP1 variants of group A were also associated with severe malaria in a study with Brazilian children. The dominant DBL1α transcripts were determined by their amplification frequencies upon reverse transcription-PCR using degenerate primers (17). However, it has been shown to be difficult to reliably reproduce results using this approach (10). Bull et al. (4) used a similar approach to study *var* gene expression in 12 clinical isolates from Tanzania. Although they were able to identify unique short DBLα sequence markers for *var* group A that correlated with the formation of rosettes, no association between *var* group expression and disease outcome was found. Recently, Kaestli et al. (16) analyzed parasites from Papua New Guinean children with asymptomatic, uncomplicated, or severe malaria infections by using quantitative real-time PCR to investigate changes in the proportion of *var* A, B, or C transcripts, using primers targeting the corresponding upstream regions. This study showed a significant increase in proportions of *var* group B transcripts in clinical cases, whereas *var* group C transcript levels were increased in asymptomatic cases. No particular involvement of *var* group A was reported.

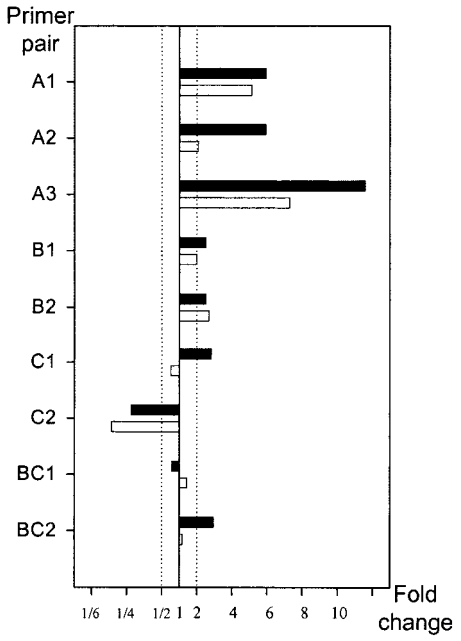


FIG. 1. Differences in *var* transcript abundance (*x*-fold changes) between $3D7_{UM}$ and antibody-selected $3D7_{SM}$ parasites (15). Transcript abundance was measured by using primers targeting *var* gene groups (white) or primers targeting single *var* genes (black) and are summarized corresponding to *var* gene groups. A twofold change in *var* transcript abundance (dashed lines) was arbitrarily defined as the cutoff for biologically significant changes in *var* transcript abundance.

The above-mentioned primers and probes were also applied in the present study, but to allow for relative comparisons to internal control genes, a new set of primers with specificity for the three main groups (A, B, and C) and *var1-3* was designed. By using these primers, *var* transcript levels were measured in cDNAs from 106 Tanzanian children with asymptomatic, uncomplicated, or severe *P. falciparum* infections.

Similar to the results of the study by Kaestli et al. (16), a larger transcript abundance of *var* group B genes was observed with an increasing degree of disease severity. Importantly, a similar pattern of transcript abundance was found for the *var* genes of group A (including *var3*). Conversely, *var* genes of group C were found to be transcribed at the same level in all sample groups. The differential transcript abundance patterns determined with the BC1 and BC2 primers corresponded to the transcript abundance patterns measured by B and C group-specific primers, respectively. These conclusions were supported by data generated by two quantitative PCR methods with primers targeting both the 5′ and 3′ ends of the genes. Although we cannot exclude the possibility that the difference in PfEMP1 expression patterns in asymptomatic children and younger symptomatic children is due to the marked age difference or the severity of infection, the data suggest that *var* group C genes are not involved in severe childhood malaria. The fact that no *var* gene group was detected at higher transcription levels in AM and UM than in SM samples is puzzling. This could be explained in several ways. Firstly, the categorization of patients into AM, UM, or SM is operational, and other host and parasite factors, including other variant surface antigens, might play equally important roles in disease outcome. Secondly, since the current knowledge of the global *var* sequence repertoire is limited, unknown transcripts not targeted by our primers may be excluded from the analysis. Finally, there is a possibility that the present primers result in biased amplification of a subset of predicted target genes. In particular, the last case might be true for the *var* group C primer pairs C2 and BC2, as indicated by the relatively large differences between estimated and observed C_T values for the 3D7 genomic DNA amplifications. Alternatively, group A or B *var* genes might be expressed in relatively larger abundance than *var* group C genes. This explanation would assume that not only the type of adhesion ligand, but also the amount of ligand, determines the adhesion phenotype. Future studies will test this hypothesis.

In logistic regression models, there was a statistically significant association between the risk of developing severe malaria and the transcript abundance of group A or B *var* genes. Thus, a twofold increase in *var* group A or B transcript abundance was associated with an increase of 20 to 61% in the risk of developing severe malaria. This observation does not indicate any difference between *var* group A and B transcript abundances in relation to disease severity. However, some indications of functional differences in group A and B *var* genes may be found in the linear regression models of *var* transcript abundance and clinical features. The association of increased parasitemia with *var* group B transcript abundance suggests that the parasites expressing these genes caused the severe

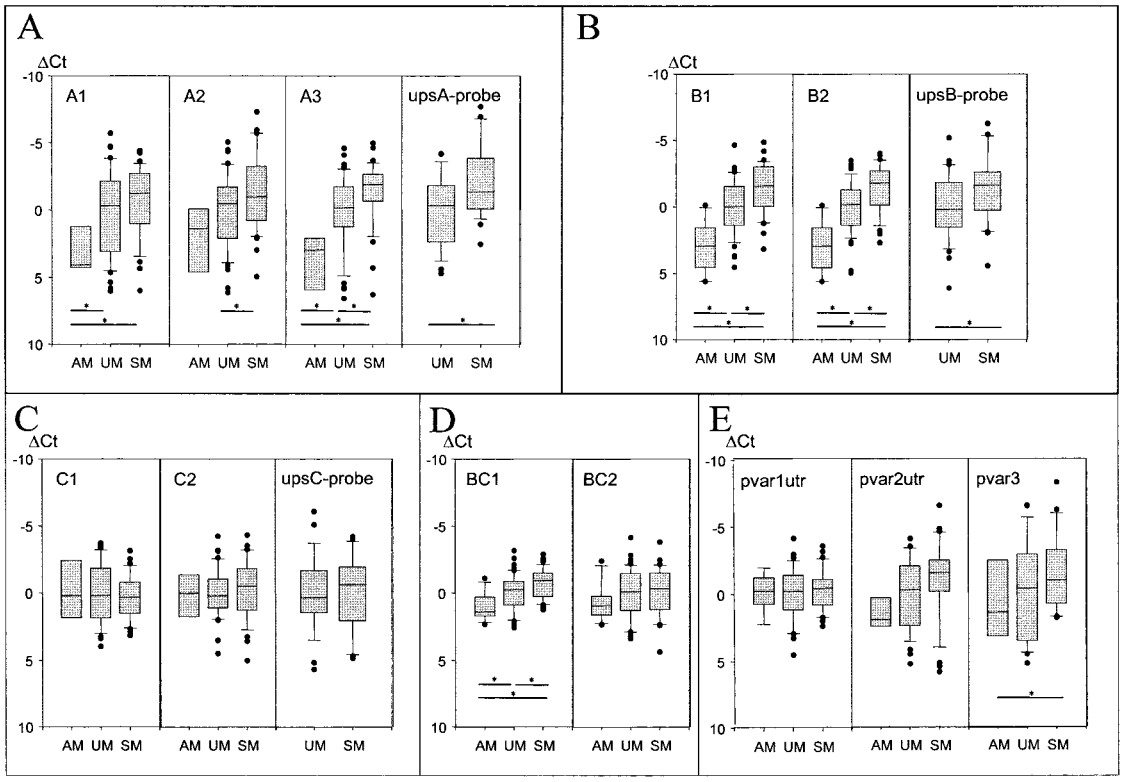


FIG. 2. Transcript abundances of *var* gene groups in parasites from children with AM and from children suffering from UM or SM. Transcript abundances are shown relative to the average abundances in uncomplicated cases (ΔC_T values). Panel A shows *var* group A transcript abundances measured with primers A1-3 in quantitative PCR and with upsA-probe in the MGB assay. Similarly, panels B and C show the transcript abundances of group B and C *var* genes, respectively. Panel D shows the transcript abundances measured with primers BC1 and BC2, targeting group B and C genes. Panel E shows transcript abundances measured with primers targeting the conserved *var* genes *var1-3*. Boxes outline 25th to 75th percentiles, with medians indicated as a line inside each box and whiskers illustrating the 5th and 95th percentiles. Horizontal lines with asterisks below the plots indicate statistically significant differences in transcript abundance between groups (one-way analysis of variance; $P < 0.05$ after Bonferroni correction).

infections in these children. In contrast, the lack of association between *var* group C transcription and parasitemia or clinical presentation supports a previous finding which suggested that group C *var* genes may be involved in establishing chronic infections (16). Similar to the case for *var* group B transcripts, a positive correlation between *var* group A transcript abundance and parasitemia would be expected in these data and in previous findings indicating the involvement of group A *var*

TABLE 3. Clinical and parasitological details of subjects

Parameter	Value for indicated group[a]			Significant relationship(s) between groups (*P*)[b]
	AM (n = 12)	UM (n = 52)	SM (n = 42)	
Mean age (mo)	46 (38, 54)	30 (26, 34)	28 (24, 32)	AM > UM/SM (<0.001)*
Sex (no. of males/no. of females)	6/6	22/30	18/24	
Mean parasitemia (parasites/200 leukocytes)	336 (144, 527)	867 (1,067, 2,206)	2,105 (1,883, 3,230)	AM < UM < SM (<0.001)*
Mean PCV (%)	29.28 (25.60, 32.95)	27.98 (26.32, 29.64)	25.37 (23.34, 27.4)	AM > UM/SM (<0.001)*
Mean lactate (mmol/liter)	3.82 (2.79, 4.84)	2.67 (2.4, 2.9)	2.9 (2.6, 2.3)	AM > UM (0.02)*
Mean glucose level (mmol/liter) (95% CI)	4.97 (4.22, 5.71)	5.1 (4.62, 5.58)	5.24 (4.68, 5.81)	0.491*
No. of days of illness	Not applicable	2.6 (2.3, 2.9)	2.8 (2.5, 3.1)	0.526**
Mean MUAC (cm)	Not determined	15.96 (15.02, 15.71)	15.36 (15.56, 16.36)	0.01**
No. of patients with prostration/ total no. of patients	0/12		40/42	
No. of patients with impaired consciousness, coma, or neurological alterations (Blantyre score of ≤3)/total no. of patients	0/12	0/52	7/42	

[a] Values in parentheses are 95% confidence intervals.
[b] *, analysis of variance/Kruskal-Wallis test; **, *t* test.

TABLE 4. Comparison of *var* group transcript abundances in parasites from patients with SM or UM

Primer pair	Fold change[a] (by $\Delta\Delta C_T$ method)	Confidence interval	*P* value (*t* test)
A1	1.6	0.6, 3.9	0.3067
A2	2.6	1.2, 5.6	0.0175
A3	2.6	1.2, 5.6	0.0148
UpsA-probe	4.3	1.6, 11.5	0.0051
B1	2.7	1.5, 4.9	0.0014
B2	2.5	1.4, 4.5	0.0020
UpsB-probe	2.8	1.2, 6.5	0.0208
C1	0.8	0.4, 1.5	0.4779
C2	1.1	0.6, 2.1	0.6836
UpsC-probe	1.1	0.4, 3.1	0.8494
BC1	1.7	1.2, 2.4	0.0050
BC2	1.1	0.6, 1.8	0.7928
Pvar1utr	1.1	0.5, 2.7	0.4375
Pvar2utr	2.0	0.8, 4.8	0.1383
Pvar3coding	2.6	0.7, 10.2	0.1611

[a] Transcript abundance in SM parasites/transcript abundance in UM parasites.

TABLE 5. Logistic regression model showing the risk of severe malaria for a twofold increase in transcript abundance of specific *var* gene groups after correcting for the effects of age, MUAC, parasitemia, and sex

Primer pair	Odds ratio	95% Confidence interval	*P*
A2	1.28	1.06, 1.56	0.012
A3	1.37	1.08, 1.75	0.010
upsA	1.35	1.02, 1.79	0.037
B1	1.47	1.08, 2.00	0.014
B2	1.52	1.14, 2.04	0.006
BC1	1.61	1.06, 2.44	0.023
upsB	1.20	0.93, 1.56	0.170

genes in severe malaria (3, 15, 17). The lack of such an association might be explained if group A PfEMP1 variants confer the strongest cytoadhesion in naïve individuals only and if group A variants only dominate in first malaria infections. Since the average age of the children enrolled in this study was 29 months, most children would have undergone several, sometimes severe, infections and would have developed some immunity against PfEMP1 variants of group A. This is supported by a trend towards a lower *var* group A transcript abundance with increasing age in UM cases.

The observed trend of more abundant *var* group A transcripts in cerebral malaria cases than in all other cases might indicate that these *var* genes play a specific role in cerebral malaria. *var2csa* has been identified as the main chondroitin sulfate A binding ligand in pregnancy-associated malaria (28). The difference seen in *var2csa* transcript abundance between the disease groups was therefore unexpected. However, for all samples, the *var2csa* transcript abundance was >100-fold lower (data not shown) than that reported for placental parasites (35) or parasites selected in vitro on CSA (29). In addition, while *var2csa* transcription appears to be controlled by similar mechanisms to those controlling group A *var* genes (11), the translation of *var2csa* transcripts, unlike that of other *var* transcripts, seems to be controlled by translation of an upstream open reading frame (22; K. W. Deitsch, personal communication). Thus, *var2csa* is most likely not responsible for the disease outcomes of these children.

var1-like genes are unique among *var* genes because they are highly conserved between parasite genomes and appear to be controlled by a unique 5′ region (36), which might indicate a specialized function for *var1* products similar to the *var2csa* gene in pregnancy-associated malaria. However, the similar abundances of *var1* transcripts in all three groups gave no indications of the function of *var1* products. This, together with the observed constitutive *var1* transcription throughout the intraerythrocytic stages in isogenic but phenotypically different parasite lines (20, 21, 29), leaves the function of *var1* molecules enigmatized.

In conclusion, the data presented here show an association between disease outcomes and the transcription of *var* subtypes in African areas where malaria is endemic. Of specific importance, the association between severe malaria in young children and *var* group A and B transcription is demonstrated and supports the notion that a vaccine based on selected PfEMP1 molecules might be feasible.

ACKNOWLEDGMENTS

We express our thanks to Boniface Idindili and Wilbert Manyilizu for the recruitment and assessment of patients, Irene Kasiga for clinical supervision, and the staffs of the Saint Francis Designated District Hospital Ifakara and the Ifakara Health Research and Development Centre for their field assistance. We highly appreciate the cooperation of the parents and children who were willing to participate in this study. The PlasmoDB database (http://www.plasmodb.org) has been a valuable resource for this work, and the database developers and researchers who made their data available are thanked.

This study was supported by the Swiss National Science Foundation (grant number 3100 AO 104043/1) and by grants from the Danish Medical Research Council (no. 22-02-0220) and the Commission of the European Communities (grant no. QLK2-CT-1999-01293) (EUROMALVAC).

REFERENCES

1. **Baruch, D. I., B. L. Pasloske, H. B. Singh, X. Bi, X. C. Ma, M. Feldman, T. F. Taraschi, and R. J. Howard.** 1995. Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell **82:**77–87.
2. **Berendt, A. R., D. J. Ferguson, and C. I. Newbold.** 1990. Sequestration in Plasmodium falciparum malaria: sticky cells and sticky problems. Parasitol. Today **6:**247–254.
3. **Bian, Z., and G. Wang.** 2000. Antigenic variation and cytoadherence of PfEMP1 of Plasmodium falciparum-infected erythrocyte from malaria patients. Chin. Med. J. **113:**981–984.
4. **Bull, P. C., M. Berriman, S. Kyes, M. A. Quail, N. Hall, M. M. Kortok, K. Marsh, and C. I. Newbold.** 2005. Plasmodium falciparum variant surface antigen expression patterns during malaria. PLoS Pathog. **18:**1.
5. **Bull, P. C., M. Kortok, O. Kai, F. Ndungu, A. Ross, B. S. Lowe, C. I. Newbold, and K. Marsh.** 2000. Plasmodium falciparum-infected erythrocytes: agglutination by diverse Kenyan plasma is associated with severe disease and young host age. J. Infect. Dis. **182:**252–259.
6. **Bull, P. C., B. S. Lowe, M. Kortok, and K. Marsh.** 1999. Antibody recognition of *Plasmodium falciparum* erythrocyte surface antigens in Kenya: evidence for rare and prevalent variants. Infect. Immun. **67:**733–739.
7. **Bull, P. C., B. S. Lowe, M. Kortok, C. S. Molyneux, C. I. Newbold, and K. Marsh.** 1998. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat. Med. **4:**358–360.
8. **David, P. H., M. Hommel, L. H. Miller, I. J. Udeinya, and L. D. Oligino.** 1983. Parasite sequestration in Plasmodium falciparum malaria: spleen and antibody modulation of cytoadherence of infected erythrocytes. Proc. Natl. Acad. Sci. USA **80:**5075–5079.
9. **Dodoo, D., T. Staalsoe, H. Giha, J. A. Kurtzhals, B. D. Akanmori, K. Koram, S. Dunyo, F. K. Nkrumah, L. Hviid, and T. G. Theander.** 2001. Antibodies to variant antigens on the surfaces of infected erythrocytes are associated with protection from malaria in Ghanaian children. Infect. Immun. **69:**3713–3718.
10. **Duffy, M. F., J. C. Reeder, and G. V. Brown.** 2003. Regulation of antigenic variation in Plasmodium falciparum: censoring freedom of expression? Trends Parasitol. **19:**121–124.
11. **Freitas-Junior, L. H., R. Hernandez-Rivas, S. A. Ralph, D. Montiel-Condado,**

O. K. Ruvalcaba-Salazar, A. P. Rojas-Meza, L. Mancio-Silva, R. J. Leal-Silvestre, A. M. Gontijo, S. Shorte, and A. Scherf. 2005. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. Cell **121:**25–36.

12. **Gardner, M. J., N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M. S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, G. I. McFadden, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser, and B. Barrell.** 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature **419:**498–511.
13. **Giha, H. A., T. Staalsoe, D. Dodoo, C. Roper, G. M. Satti, D. E. Arnot, L. Hviid, and T. G. Theander.** 2000. Antibodies to variable Plasmodium falciparum-infected erythrocyte surface antigens are associated with protection from novel malaria infections. Immunol. Lett. **71:**117–126.
14. **Howard, R. J., and J. W. Barnwell.** 1984. Roles of surface antigens on malaria-infected red blood cells in evasion of immunity. Contemp. Top. Immunobiol. **12:**127–200.
15. **Jensen, A. T., P. Magistrado, S. Sharp, L. Joergensen, T. Lavstsen, A. Chiucchiuini, A. Salanti, L. S. Vestergaard, J. P. Lusingu, R. Hermsen, R. Sauerwein, J. Christensen, M. A. Nielsen, L. Hviid, C. Sutherland, T. Staalsoe, and T. G. Theander.** 2004. Plasmodium falciparum associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A var genes. J. Exp. Med. **199:**1179–1190.
16. **Kaestli, K., I. A. Cockburn, A. Cortés, K. Baea, J. A. Rowe, and H.-P. Beck.** Virulence of malaria is associated with differential expression of *Plasmodium falciparum var* gene subgroups in a case-control study. J. Infect. Dis., in press.
17. **Kirchgatter, K., and H. A. Portillo.** 2002. Association of severe noncerebral Plasmodium falciparum malaria in Brazil with expressed PfEMP1 DBL1 alpha sequences lacking cysteine residues. Mol. Med. **8:**16–23.
18. **Kraemer, S. M., and J. D. Smith.** 2003. Evidence for the importance of genetic structuring to the structural and functional specialization of the Plasmodium falciparum var gene family. Mol. Microbiol. **50:**1527–1538.
19. **Kyes, S., P. Horrocks, and C. Newbold.** 2001. Antigenic variation at the infected red cell surface in malaria. Annu. Rev. Microbiol. **55:**673–707.
20. **Kyes, S. A., Z. Christodoulou, A. Raza, P. Horrocks, R. Pinches, J. A. Rowe, and C. I. Newbold.** 2003. A well-conserved Plasmodium falciparum var gene shows an unusual stage-specific transcript pattern. Mol. Microbiol. **48:**1339–1348.
21. **Lavstsen, T., P. Magistrado, C. C. Hermsen, A. Salanti, A. T. Jensen, R. Sauerwein, L. Hviid, T. G. Theander, and T. Staalsoe.** 2005. Expression of Plasmodium falciparum erythrocyte membrane protein 1 in experimentally infected humans. Malar. J. **4:**21.
22. **Lavstsen, T., A. Salanti, A. T. Jensen, D. E. Arnot, and T. G. Theander.** 2003. Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions. Malar. J. **2:**27.
23. **Nielsen, M. A., T. Staalsoe, J. A. Kurtzhals, B. Q. Goka, D. Dodoo, M. Alifrangis, T. G. Theander, B. D. Akanmori, and L. Hviid.** 2002. Plasmodium falciparum variant surface antigen expression varies between isolates causing severe and nonsevere malaria and is modified by acquired immunity. J. Immunol. **168:**3444–3450.
24. **Nielsen, M. A., L. S. Vestergaard, J. Lusingu, J. A. Kurtzhals, H. A. Giha, B. Grevstad, B. Q. Goka, M. M. Lemnge, J. B. Jensen, B. D. Akanmori, T. G. Theander, T. Staalsoe, and L. Hviid.** 2004. Geographical and temporal conservation of antibody recognition of *Plasmodium falciparum* variant surface antigens. Infect. Immun. **72:**3531–3535.
25. **Pain, A., D. J. Ferguson, O. Kai, B. C. Urban, B. Lowe, K. Marsh, and D. J. Roberts.** 2001. Platelet-mediated clumping of Plasmodium falciparum-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. Proc. Natl. Acad. Sci. USA **98:**1805–1810.
26. **Robinson, B. A., T. L. Welch, and J. D. Smith.** 2003. Widespread functional specialization of Plasmodium falciparum erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. Mol. Microbiol. **47:**1265–1278.
27. **Rowe, A., J. Obeiro, C. I. Newbold, and K. Marsh.** 1995. *Plasmodium falciparum* rosetting is associated with malaria severity in Kenya. Infect. Immun. **63:**2323–2326.
28. **Salanti, A., M. Dahlback, L. Turner, M. A. Nielsen, L. Barfod, P. Magistrado, A. T. Jensen, T. Lavstsen, M. F. Ofori, K. Marsh, L. Hviid, and T. G. Theander.** 2004. Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. J. Exp. Med. **200:**1197–1203.
29. **Salanti, A., T. Staalsoe, T. Lavstsen, A. T. Jensen, M. P. Sowa, D. E. Arnot, L. Hviid, and T. G. Theander.** 2003. Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering Plasmodium falciparum involved in pregnancy-associated malaria. Mol. Microbiol. **49:**179–191.
30. **Schellenberg, D., C. Menendez, J. Aponte, C. Guinovart, H. Mshinda, M. Tanner, and P. Alonso.** 2004. The changing epidemiology of malaria in Ifakara Town, southern Tanzania. Trop. Med. Int. Health **9:**68–76.
31. **Smith, J. D., C. E. Chitnis, A. G. Craig, D. J. Roberts, D. E. Hudson-Taylor, D. S. Peterson, R. Pinches, C. I. Newbold, and L. H. Miller.** 1995. Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell **82:**101–110.
32. **Snow, R. W., J. A. Omumbo, B. Lowe, C. S. Molyneux, J. O. Obiero, A. Palmer, M. W. Weber, M. Pinder, B. Nahlen, C. Obonyo, C. Newbold, S. Gupta, and K. Marsh.** 1997. Relation between severe malaria morbidity in children and level of Plasmodium falciparum transmission in Africa. Lancet **349:**1650–1654.
33. **Staalsoe, T., M. A. Nielsen, L. S. Vestergaard, A. T. Jensen, T. G. Theander, and L. Hviid.** 2003. In vitro selection of Plasmodium falciparum 3D7 for expression of variant surface antigens associated with severe malaria in African children. Parasite Immunol. **25:**421–427.
34. **Su, X. Z., V. M. Heatwole, S. P. Wertheimer, F. Guinet, J. A. Herrfeldt, D. S. Peterson, J. A. Ravetch, and T. E. Wellems.** 1995. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. Cell **82:**89–100.
35. **Tuikue Ndam, N. G., A. Salanti, G. Bertin, M. Dahlback, N. Fievet, L. Turner, A. Gaye, T. Theander, and P. Deloron.** 2005. High level of var2csa transcription by Plasmodium falciparum isolated from the placenta. J. Infect. Dis. **192:**331–335.
36. **Vazquez-Macias, A., P. Martinez-Cruz, M. C. Castaneda-Patlan, C. Scheidig, J. Gysin, A. Scherf, and R. Hernandez-Rivas.** 2002. A distinct 5′ flanking var gene region regulates Plasmodium falciparum variant erythrocyte surface antigen expression in placental malaria. Mol. Microbiol. **45:**155–167.
37. **Voss, T. S., J. K. Thompson, J. Waterkeyn, I. Felger, N. Weiss, A. F. Cowman, and H. P. Beck.** 2000. Genomic distribution and functional characterisation of two distinct and conserved Plasmodium falciparum var gene 5′ flanking sequences. Mol. Biochem. Parasitol. **107:**103–115.
38. **World Health Organization, Communicable Diseases Cluster.** 2000. Severe falciparum malaria. Trans. R. Soc. Trop. Med. Hyg. **94**(Suppl. 1)**:**S1–S90.

Editor: W. A. Petri, Jr.

Chapter Four

Genetic Diversity of Expressed *Plasmodium falciparum var* genes from Tanzanian children with Severe Malaria

Chapter Four

Genetic Diversity of Expressed *Plasmodium falciparum var* genes from Tanzanian children with Severe Malaria

Joseph Paschal Mugasa[1], Weihong Qi[2], Sebastian Rusch[2], Matthias Rottman[2] and Hans-Peter Beck[2*]

[1]*Ifakara Health Research and Development Centre, P.O.BOX 53, Ifakara, Morogoro, Tanzania*

[2]*Swiss Tropical Institute, Socinstrasse 57, Postfach 4002 Basel, Switzerland*

*Corresponding address
Swiss Tropical Institute
Socinstrasse 57
CH 4002 Basel
Tel.: +41-61-284 8116, Fax: +41-61-284 8101
E-mail address: hans-peter.beck@unibas.ch (HP Beck)

This article will be submitted to the International Journal of Parasitology

Abstract

Severe malaria has been attributed to the expression of a restricted subset of the *var* multi-gene family, which encodes for *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1). PfEMP1 mediates cytoadherence to a variety of host cell receptors and cause sequestration of infected erythrocytes in post capillary venules of the vital organs such as the brain or placenta which is a key element in the pathology of malaria. *var* genes are highly diverse, and can be classified in three major groups (upsA, B and C) and two intermediate groups (B/A and B/C) based on the genomic location, gene orientation and the upstream sequences. We have shown previously that *var* group A and B are up-regulated in clinical malaria compared to children with asymptomatic infections. We studied subsequently the genetic diversity of expressed *var* genes associated with severe childhood malaria. By use of biotinylated magnetic beads tagged the reverse complement of the conserved exon 2, full-length *var* mRNA was isolated and reverse transcribed into *var* cDNA. Different N-terminal domain tags were amplified by PCR, cloned and sequenced from children isolates with severe (SM) and asymptomatic malaria (AM). Our analyses shows high sequence diversity of the amplified *var* DBL-1α and upstream regions with minimal overlap among the isolates, providing strong evidence that *var* gene repertoire is immense and indefinite in high endemic areas. *var* DBL-1α sequences from AM isolates were more diverse with more singletons ($p<0.05$) than those from SM cases. Furthermore, few *var* DBL-1α sequences from SM patients were rare and restricted suggesting that certain PfEMP1 variants are predisposed for inducing severe infection.

Keywords: *Plasmodium falciparum*; severe malaria; *var* genes; PfEMP1; Repertoire; Diversity.

1. Introduction

Plasmodium falciparum malaria infections continue to be the primary cause of morbidity and mortality in many developing countries, with an estimated 1.5 -2.7 million deaths annually (Breman & Holloway 2007). The burden of disease is greatest in children < 5 years old where, much of the mortality is attributable to severe malaria. Despite extensive research efforts there is no vaccine to date. Vaccine development is encouraged by the fact that children living in endemic areas attain conditional immunity to severe malaria after a relatively few number of episodes during childhood. However, the genetic diversity between different parasites isolates poses an obstacle for vaccine development.

Severe malaria, a life threatening form of the disease, is believed to be mediated by cytoadhesion of *P. falciparum*- infected erythrocytes to a variety of host cell receptors on the endothelial cells of the host. This causes sequestration into vital organs such as post-capillary venules of the brain, kidneys, lungs or placenta (Miller et al 2002). *Plasmodium falciparum* expresses a number of proteins on the surface of infected erythrocytes that play a key role as both virulence factors in malaria pathology by mediating cytoadherence, and as targets for naturally acquired immunity (Bull & Marsh 2002, Kyes et al 2001). The virulence of *P. falciparum* has been associated with the expression of *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) on the surface of infected erythrocytes. PfEMP1 is large 200-400 kDa, highly polymorphic antigen, encoded by a family of ~ 60 *var* genes per haploid genome (Gardner et al 2002). The *var* genes present a two-exon structure encoding a conserved C terminus that contains a predicted transmembrane region and a polymorphic extracellular N terminus. This part is composed of a number of cysteine-rich domains that are involved in sequestration of the parasite in the microvasculature (Baruch et al 1995, Smith et al 1995, Su et al

1995). The duffy binding like-domain α (DBLα) is the most conserved domain within the *var* gene domains. It is located in the N-terminal head structure of PfEMP1, and was found to mediate rosetting and cytoadherence (Chen et al 1998, Rowe et al 1997). A range of host receptors appear to be involved in binding such as heparin, heparin sulfate, complement receptor 1 (CR1) and blood group A antigen (Chen et al 1998, Mayor et al 2005, Vogt et al 2003). Rosetting and endothelial binding are closely associated with severe falciparum malaria (Fairhurst et al 2005, Rowe et al 2000).

var genes have been classified into three major groups (A, B and C) and two intermediate groups (B/A and B/C) based on the presence of one of the three conserved 5' upstream sequences (UpsA, B or C), position and orientation of the gene within a chromosome (Lavstsen et al 2003, Voss et al 2000). Severe malaria has been associated with expression of a restricted and antigenically semi conserved subset of variant erythrocyte surface antigens (Bull et al 2000, Nielsen et al 2002). Of the different malaria diseases syndromes, the role of PfEMP1 protein is best understood for pregnancy-associated malaria (PAM) Gamain et al (2007).

Few studies have investigated the expression of *var* genes in field isolates representing different forms of severe malaria (Ariey et al 2001, Bull et al 2005, Kaestli et al 2004, Kirchgatter & del Portillo 2002, Kyriacou et al 2006). These studies suggested that the transcription patterns of *var* genes vary between different malaria manifestations. Differences in epidemiology, severe disease classification, and *var* classification have also made comparison between studies difficult. Using quantitative real time reverse transcription PCR (qRT-PCR), we have previously shown that group A and B *var* transcripts were up-regulated in children with clinical malaria as opposed to asymptomatic infections from

Tanzania (Rottmann et al 2006) qRT- PCR is a standard method for detection and quantification of genes expression levels (Bustin 2000), however, this technique is not informative in studying genetic diversity of genes. Given the proposed importance of immunity to PfEMP1 in protection against malaria, it is essential that we gain a better understanding of diversity of this molecule at the sequence level and how such diversity influences the development of protective immunity. In this study, we examined the diversity of expressed PfEMP1 repertoires in parasite populations isolated from children with severe malaria.

2. Materials and Methods

2.1 Sample collection

Samples used in this study were collected in a severe malaria (SM) case control study that has been described in details previously (Rottmann et al 2006). Briefly, children aged <59 months admitted with severe malaria at Saint Francis Designated District Hospital (StFDDH), Ifakara, Tanzania were recruited into the study after the informed consent was obtained from children's parents or guardians. One to two ml of venous blood from each child was drawn into an EDTA tube (Vacutainer, Becton Dickinson, Rutherford, NJ, USA). Between 50 – 100 μL of this whole blood was then mixed with 2 volumes of 6 M guanidine HCl, 50 mMTris pH 8.0, 20mM EDTA and kept at -20 °C for gDNA isolation. The remaining RBCs were separated from serum by centrifugation and washed with 40 ml phosphate buffered saline, 5 volume of TRIzol reagent (Invitrogen) were added to the RBC pellet before preservation at -70 °C until later use.

Exclusion criteria were confirmed co-infections, malnutrition (mid -upper arm circumference [MUAC] of ≤ 12 cm), haemoglobin ≤ 5 g/dL, lactate ≥ 5 mmol/L, glucose, ≤ 2.2 mmol/L) or antimalarial treatment during the last 14 days. A total of 52 patients with SM were recruited. Eight patients met the inclusion criteria and were grouped into cerebral malaria according to WHO guidelines (WHO 2000) and the modified Blantyre coma score ≤ 3 (Molyneux et al 1989). In nearby villages children of the control group were recruited with asymptomatic malaria (AM) *i.e* presence of *P. falciparum*, axillary temperature of ≤ 37.5 °C and no other symptoms. Informed consent was sought from the children's parents or guardians, children aged < 59 months were screened for *P. falciparum* infection by using a rapid diagnostic test (RDT), Paracheck® Pf (Orchid Biomedical Systems, Goa, India). Participating children who were found positive by RDT later had their infection

status confirmed microscopically by Giemsa-stained thick and thin blood films at IHRDC laboratory. A total of 19 children with AM were recruited. Ethical clearance for this study was obtained from the Ifakara Health Research and Development Centre and the Medical Research Coordinating Committee of the National Institute for Medical Research in Tanzania.

2.2 DNA extraction and genotyping

Genomic DNA was extracted from frozen blood using Qiamp kit (Qiagen) by following the manufacturer's instructions. The minimum number of genotypes per isolate was evaluated by Genescan as described by Falk et al (2006) and confirmed by genotyping PCR with MSP2 primers as described by Felger et al (1993), briefly, 1 μL of purified genomic DNA was used in a 20 μL primary PCR reaction, followed by restriction fragment length polymorphism (RFLP) of the nested PCR product to record the number of infecting strain per isolate.

2.3 Isolation of full-length *var* transcripts and RT- PCR

For total RNA extraction blood pellets in TRIzol were thawed on ice. TRIzol (Invitrogen) was pre-warmed to 37 °C and used by following the manufacturer's protocol. Extraction with TRIzol was performed twice, to decrease DNA contamination. RNA was treated with 3 U of RQ1 RNase-Free DNase (Promega), followed by the second extraction with TRIzol. Full-length *var* mRNA with an acidic terminal sequence (ATS) was isolated by using magnetic beads tagged with an anti-ATS oligonucleotide as previously described with modifications (Kaestli et al 2004). Briefly, RNA was dissolved in (5 mmol/L Tris/0.5 mmol/L EDTA), and mixed with binding buffer (0.5 mol/L LiCl, 1 mmol/L EDTA, 10 mmol/LTris, pH 7.5), 15 mmol/L DTT, 40 U RNase OUT (Invitrogen) and 1 pmol of biotinylated

oligonucleotide complementary to the conserved sequence in ATS domain (Biotin-5'- GGTTC(A/T)A(A/G)TAC(C/T)ACTTC(A/T)AT(C/T)CCTGGT(A/G)CAT

ATATATCATTAATATCCAATTCTTCATA(C/T)TCACTTC(T/G)GA

(A/T/G)GA-3') incubated at a temperature gradient from 65 °C to 4 °C over 30 minutes. Meanwhile 150 mg of Dynabeads M-280 streptavidin (Dynal Biotech, ASA, Oslo, Norway) was washed as suggested by manufacturer, dissolved in 5.5 mmol/L LiCl and added to the beads-ATS-mRNA hybrid. The mixture was uniformly mixed by rotating for 30 min at 37 °C. The biotinylated beads-ATS-mRNA complex was washed three times with washing buffer (10 mmol/L Tris, 1 mmol/L EDTA, 0.15 mol/L NaCl, pH 7.5) and 1 time with 10 mmol/L Tris. Reverse transcription (RT) into single stranded cDNA was performed on captured mRNA, primed by random hexamere oligonucleotides in the concentration of 300 ng (Invitrogen) using Sensiscript (Qiagen), reverse transcriptase, following the manufacturer's instructions in a final volume of 20 μL. A second RNA aliquot was equally treated except that reverse transcriptase was omitted in the RT-step; this sample served as a control for proving the absence of gDNA. After RT, cDNA was treated with RNaseA (Promega) and 1 μL was used for various PCRs.

2.4 Amplification of *var* sequences

The DBL-1α domain of *var* genes was amplified from 1 μL cDNA template by PCR with the Advantage cDNA polymerase mix (CLONTECH) using primer sets shown in Table 1 to generate a PCR product of about ~ 400- 500 bp in length. PCR was performed in 25 μL 1.5 mM $MgCl_2$, 200 μM dNTP mix, 1 μM each primer. The cycling conditions were 30 cycles of 94 °C for 30 s, 1 min at the annealing temperature specified in (Table 1) and for 70 s at 68 °C. To study the genetic diversity of the untranslated upstream sequences, three different PCRs were carried out on cDNA using designed degenerate primers based upon sequence alignments of 3D7 *var* genes. Primers were designed to amplify the homology

blocks in upstream sequences (5' UTR, 200bp region on upsA, 400 bp and 440 bp for upsB and upsC, respectively) and the reverse primer was chosen from homology block H of the first DBL-1α domain (Kraemer et al 2003). The primers were tested on genomic DNA from the 3D7 isolate. Amplification of DBLα-CIDRβ fragments was done by using primers shown in Table 1. Controls minus reverse transcriptase were amplified in parallel in each reaction to control for gDNA contamination. If a PCR product was obtained in the control, the positive sample was excluded from the analysis.

2.5 Cloning and sequencing

An aliquot of 5 μL of each PCR product was visualized in a 1% agarose gel, the remaining PCR product was purified by using the NucleoSpin® Extract II (Macherey & Nagel). The eluted DBL-1α fragments were cloned into the pGEM-T vector (Promega) according to the manufacturer's instructions and transfected into *Escherichia coli* SURE cells (Stratagene). Untraslated upstream regions and DBL-1α-CIDR1 products were cloned into the pCR®4-TOPO (Invitrogen) and transformed into TOP10 cells. This vector was more efficient for large fragments (> 1 kb) and low concentrated PCR products (< 5 ng/μL). From each clinical isolate an average of 50 white colonies that were found to be positive by PCR screening were processed for sequencing (Perfectprep® Plasmid 96 Vac Direct Bind kit, Eppendorf). The size of the insert was checked in purified plasmids using restriction enzymes NotI and NcoI (New England BioLab) for pGEMTs plasmids. EcoRI (New England BioLab) was used for TOPO following manufacturer's instructions. Sequencing was carried out using the T7 and SP6 primers for pGEM-T vector, whereas M13 forward and reverse primers for the pCR®4-TOPO using 96 capillary automated sequencing systems 3700 (Applied Biosystems). A multiple-sequence alignment of sequences derived from the same clinical isolate was carried out to allow the exclusion of PCR derived mutations. Two sequences were considered to be identical when ≥ 96 % amino acid sequence identity was detected.

2.6 Sequence analysis

DNA sequences were assembled and analyzed using ContigExpress in the Vector NTI Advance™ 10 software (Invitrogen) and BLAST from the National Centre for Biotechnology information webpage: (http://www.ncbi.nlm.nih.gov/BLAST/). BLAST analysis against the 3D7 genome database was performed at *P. falciparum* GeneDB webpage: (http://www.genedb.org/genedb/malaria/) and PlasmoDB 5.4 (https://www.pasmoDB.org). The DNA sequences were translated using RevTrans1.4 (Wernersson & Pedersen 2003). Amino acids were aligned with either CLUSTALW 1.8 or MUSCLE (Edgar 2004) using default parameters and edited with Bioedit version 7.09 with minor manual adjustment where necessary. Sequences were further categorized into sequence types (STs) by BLASTCUT analysis (Altschul et al 1990) by which sequences sharing ≥ 96 % sequence identity were assigned the same ST.

2.7 Phylogenetic analysis

Phylogenetic analyses were conducted on multiple sequence alignment of the 3 most dominant sequences from each isolates. Because *var* genes are subject to intra-genic recombination (Freitas-Junior et al 2000), synonymous substitutions are likely saturated and DNA sequence analysis would be quite noisy in constructing phylogenetic trees (Russo et al 1996). Therefore, we used protein sequences rather than nucleotide sequences. Two methods were employed in constructing the phylogenic trees. Neighbour-Joining (NJ) trees were constructed by using MEGA 4.0 (Tamura et al 2007). The reliability of internal branches for NJ was assessed with 1000 bootstrap pseudoreplicates using 'pairwise deletion option' of amino acid sequences with p-distance. SplitsTree4 version 4.7 was used to construct the phylogenetic network (Huson & Bryant 2006) using the Neighbour-Net distances transformation and equal angle splits transformation.

3. Results

3.1 Sample collection and clinical data

A total of 15 children were used in the analysis of the present study 8 of which were found to have SM with cerebral manifestation, Blantyre score ≤ 3 and 7 children with asymptomatic *P. falciparum* malaria (12 children were dropped from analysis as they were confirmed negative for *P. falciparum* microscopically). Clinical and epidemiological assessments of isolates from SM and AM are presented in Table 2. There was a significant difference in age between the two groups AM: median 52, range 24-59, SM: median 28.5, range 14-40 (*P=0.02*, Kruskal-Wallis test). There was a highly significant differences in parasitemia between clinical categories (*P*=0.0012, Kruskal-Wallis test).

3.2 Multiplicity of infection

msp2 genotyping indicated that 87.5% (7/8 SM isolates) had multiple clone *P. falciparum* infections (2-4 clones) with an average of 2.6 infecting clones. All isolates from AM children had multiple infections (2 or 3) with an average of 2.4 infecting clones (Table 3a).

3.3 DBL-1α sequences types generated

A total of 665 *var* DBL-1α clones (~ 400-500 bp) were successfully sequenced. Of these, 305 sequences were originating from AM children (Table 3a) and were assembled into 224 Sequence types (STs) *ie* distinct DBL-1α sequences. The remaining 360 sequences from SM children were assembled into 300 STs. AM patient samples had more singletons ($p<0.05$) than SM isolates. Assembled STs showed an extreme diversity in sequence reflecting the high recombination and mutation rates in the DBL-1α domain. Multiple-sequence alignment of the DBL-1α sequences showed conserved islands of homology. The dominant sequence from each isolate was blasted against the 3D7 genome. The blasted sequence was assigned the name of the identified 3D7 gene with the high scoring segment pair (Table 3a).

3.4 Upstream sequences

Three hundred and thirty sequences were generated from *var* upstream regions and 103 sequences from DBLα-CIDRβ fragments (Table 3b). Generally, the PCR amplification and cloning efficiencies were low in some isolates. This highlights the high diversity of *var* genes in field isolates, as primers designed were based on 3D7 *var* genes and these might not amplify sequences from natural isolates. More than one distinct sequence was detected in each isolate. All isolates showed a predominant upstream sequence (Table 3b). The predominant sequence from each isolate (5′ UTR-DBL-1α or the DBL-1α fragment) was blasted against the 3D7 genome. UpsA fragments were found to have higher scoring against 3D7 *var* gene both in DBL-1α and the entire upsA type 5' UTR-DBL-1α fragment in contrast to *var* group B and C. This provides evidence that *var* group A are more conserved between field isolates and 3D7 genome.

3.5 Distribution of DBL-1α expressed sequence tags

The number of distinct transcribed DBL-1α *var* gene sequences detected per isolate varied from 8 to 25 (Table 3a). All isolates showed a predominant sequence as well as minor transcripts and unique sequence types. In cluster analysis some of the DBL-1α sequences were found to be shared among isolates (*i.e* overlapping). A number of transcripts were found in both groups (SM & AM) and others were specifically found either in SM or in AM isolates. Some sequences were unique to a particular isolate and were not found in other isolates. The distributions of the STs in our 15 isolates are shown in Figure 1. There was no significant difference in the number of distinct DBL-1α sequences per isolate detected in both clinical groups AM: median 20, range 8-25, SM: median 17.5, range 12-23, (P=0.72, Kruskall-Wallis test).

3.6 Distribution of DBL-1α expressed sequences tags in clinical isolates

All sequences generated were classified into six DBL-1α sequence tag groups by using text string software in Ms Excel and Perl which was kindly provided by Dr Pete Bull (KEMRI, Kilifi, Kenya). This classification of DBL-1α sequence tags was previously explained in detail by (Bull et al 2005). In short, it is based on counting the number of cysteine-residues within the tagged region, and in a set of sequence motifs at 4 positions of limited variability (PoLV 1-4). Figure 2 shows the distribution of PoLV/cys groups in clinical isolates. Analysis of DBLα sequences generated in the present study corresponded well with the cysteine/PoLV grouping Figure 3 shows the proportional distribution of PoLV motifs between the clinical isolates. Sequence "signature tags" and the group of the dominant sequence from each isolate are shown in Table 3. A significant association of Cys2 sequence tags (groups 1-3) with SM isolates ($p < 0.0001$, CMH test), with an odds ratio of 2.5 (95% C.I = 1.78 - 3.4) was observed. These findings support previous reports that DBL-1α sequences associated with severe disease have a reduced number of cysteines (Bull et al 2005, Kirchgatter & del Portillo 2002, Kyriacou et al 2006). A two sample test of proportion was used to test whether expressed PoLV motifs are associated with a particular disease phenotype SM or AM ($p<0.0001$). A range of expressed PoLV motifs were found strongly associated with severe malaria in PoLV1 (LDLY and MFKR), PoLV2 (FREY and LRVE), PoLV3 (NAIT and RAIT) and in PoLV4 (LTNL and PTNL).

3.7 Cumulative diversity of DBL-1α sequences in clinical isolates

To estimate the size of the *var* gene repertoire in the parasite population under study, the rate of how distinct DBL-1α sequences changed was simulated with increasing sample size. This simulation was performed separately for AM and SM, as well as the combined data (AM & SM). The empirical plots were fitted by a linear function. The curves did not plateau with the DBL-1α sequences generated from the Ifakara area (Figure 4). The repertoire of expressed *var* genes was

unlimited, thus, *var* gene diversity in this local population seems to be immense and unrestricted. However, a minimal overlap among *var* genes was found in different isolates. A similar finding has been reported Barry et al (2007) for the cumulative DBL-1α sequences from genomic DNA in the Amele, PNG, and for the global population, where in more than 1000 sequences from 59 isolates plus the entire 3D7 *var* repertoire the saturation point of the *var* gene repertoires could not be reached.

3.8 Phylogenetic analysis

To study the sequence diversity between the SM and AM groups, a phylogenetic network was constructed using the 3 most dominant DBL-1α variants expressed from each SM or AM isolate. The analyzed sequences clustered in two distinct groups. The majority of the DBL-1α isolates from severe malaria clustered together and belonged to *var* group A and B/A. AM isolates formed another cluster mainly consisting of *var* B, B/C or C (Figure 5). To further study relationships between sequences, 9 DBL-1α sequences from the 3D7 genome were incorporated, 3 from each group A, B and C in the phylogenetic tree construction. Among the sequence included was *var* group A (PF11_0008) which has been shown to be highly transcribed in the NF54 isolate (Lavstsen et al 2005) and *var* group B (PF10_0406) which has been detected previously as a major transcript in 3D7B2 and 3D7B1 samples (Peters et al 2002).

The DBL-1α sequences analyzed were found to cluster into two distinct clades. The majority of SM isolates and the 3D7 DBL-1α sequences clustering together belonged to *var* group A and B/A. AM isolates and the other remaining 3D7 DBL-1α sequences formed another cluster majority belonging to *var* group B, B/C or C (Figure 6). These findings again support the hypothesis that SM is caused by a restricted subset of *var* gene family that belongs to *var* group A or B/A whilst the

non severe form of malaria is attributed to other *var* gene groups. Similarly the phylogenetic approach by Kyriacou et al (2006) when comparing DBL-1α sequence tags from Mali, identified *var* group A and B/A to be more frequent among parasites isolated from children with cerebral malaria than those of hyperparasitemia patients.

A multiple-sequence alignment of 3 dominant upsA sequences from clinical isolates together with upsA sequences from the 3D7 genome showed the existence of short islands of homology, conserved in all isolates suggesting that they might be structurally important. Phylogenetic analysis of the three upsA dominant sequences from clinical isolates and the 3D7 upsA sequences, showed an even distribution among clinical isolates (Figure 7). Two different methods for phylogenetic tree construction were used (MEGA 4.0 and SplitsTree 4.7). Both methods yielded similar tree topologies.

4. Discussion

Studies on *var* genetic diversity are important in understanding malaria pathogenesis and feasibility for designing a disease intervention such as a vaccine or other therapeutic approach. In the present study, we examined *var* gene expression from clinical isolates of children with severe malaria and asymptomatic infections from Tanzania. Dominant expression of one particular *var* gene was found, together with less abundant variant transcripts and unique sequences in each isolate. However, the dominant sequences differed between the isolates. This suggests that each parasite contains its own form of *var* gene variants. This has the consequences that exposure to multiple infections and hence *var* gene products do not necessarily confer immunity to future malaria infections (Fowler et al 2002, Trimnell et al 2006).

By analyzing the expressed *var* genes repertoires in severe malaria cases versus asymptomatic controls, we have shown that the diversity within the *var* gene family is enormous with minimal degree of overlaps between isolates. Recently, Kyriacou et al (2006) have found a minimal overlap in *var* genes repertoires after they analyzed the expressed sequence tags from Malian children with malaria infections. However, Albretch et al (2006) reported huge overlapping *var* gene repertoires in the Western Amazon isolates. *var* repertoires of natural parasite populations found within specific geographical regions showed a degree of overlapping, suggesting the circulation of a similar *var* gene repertoire. This has important implications for the acquisition of long-term immunity by the exposed individuals (Barry et al 2007).

The diversity of *var* genes within a natural *P. falciparum* population in a particular geographical region is difficult to define, and it is also difficult to assess whether the diversity is constant due to functional constrain on this molecule, fluctuating or constantly turning over, and how fast the turnover rate of the PfEMP1 repertoires could be. Changes in the *var* repertoire are believed to be due to high allelic and ectopic recombination rates of *var* genes in field isolates (Conway et al 1999, Freitas-Junior et al 2000, Taylor et al 2000b) which are influenced by transmission intensity. The diversity of the PfEMP1 repertoire of parasites in a given geographical area is a key factor in the development of clinical immunity. The vast antigenic diversity and complexity of *var* gene repertoires in parasite populations may explain why individuals are repeatedly susceptible to *P. falciparum* infections and never develop sterilizing immunity. The antigenic variation and high switching rate of *var* gene expression are effective mechanisms adopted by *P. falciparum* to evade the host's immune system, for their survival and effective transmissions.

In our study, several sequences were observed more frequently than other sequences within individual patients. Whether this is due to the primer and cloning bias or whether it can be explained by presence of multiple infections as was the case in most of the patients remains open. Consistent with previous studies of *var* gene diversity (Bull et al 2005, Fowler et al 2002, Kirchgatter et al 2000, Kyriacou et al 2006, Taylor et al 2000b), the variability of the DBL-1α and upstream sequences within an isolate was found to be similar to different isolates in all the groups (SM &AM). Nevertheless, AM isolates were more diverse as reflected by the presence of more singletons. This suggests that *var* genes associated with asymptomatic infection have an enormous repertoire which could explain the difficulty to acquire immunity to mild or asymptomatic malaria.

Isolates from children with severe malaria were predominantly found to transcribe *var* genes with a DBL-1α domain that had a reduced number of cysteine residues which is the characteristic of *var* group A or B/A. Similar results have been reported previously from other research groups in Mali, Kenya and Brazil (Bull et al 2005, Kirchgatter & del Portillo 2002, Kyriacou et al 2006). These results highlight that severe malaria is caused by a restricted subset of *var* genes and probably *var* group A and B/A are involved in severe disease. This finding supported our previous study, in which we had shown by quantitative PCR that *var* group A was up regulated in children with cerebral malaria (Rottmann et al 2006). However, most studies on *var* gene family have been relying on the use of DBL-1α fragments (Kyes et al 1997). DBL-1α primers amplify only a small fragment of the *var* gene that is more conserved than other *var* domains and that is found in most of PfEMP1 proteins. Due to the complex nature of *var* genes it has been difficult to clone and sequence larger fragments. Larger fragments of the *var* genes would provide additional information on understanding *var* gene transcription and its association to disease phenotype.

Clustering analysis revealed several "unique sequences" of *var* gene were transcribed only in isolates from patients with severe malaria. Expression of these "unique sequences" in a patient who lacks a pre-existing antibody response against this variant might trigger the development of severe malaria. Once exposed to these virulent *var* genes individuals living in endemic areas may acquire immunity to severe malaria. In areas of high endemicity this might happen early in life after a relatively few number clinical episodes. Comparisons of PoLV motifs distribution between the clinical groups, 8 motifs were identified which were highly associated with severe disease. Recently, Normark et al (2007) identified 15 DBL-1α sequence degenerate motifs pertinent to severe disease states and 3 motifs associated with high rosetting after analysing 93 patients with well characterized disease state based on the MOTIFF algorithm. These findings also support the hypothesis that disease phenotypes are correlated with the expression of certain PfEMP1 variants and motifs, which is relevant information for vaccine development and understanding disease pathogenesis.

The distribution of PF11_0008, a group A *var* gene, which previously has been identified in the 3D7 genome and the isogenic isolate NF54 (Lavstsen et al 2005), was found in three SM isolates (ISM11, ISM33, ISM48) and in one AM sample (IAM17), although in low frequencies. This is an indication that the *var* genes of laboratory strains are shared among the field isolates.

In conclusion, we have shown that the *var* family is highly diverse in natural *P. falciparum* populations, however, the diversity was more restricted in severe malaria than in asymptomatic isolates, and this finding suggests a fundamental role played by different subset of *var* transcripts in disease syndromes. Further studies for analysis of this molecule are required from many geographical regions with well defined malaria infections. This approach might provide the basis for vaccine or chemotherapy targets. To gain better understanding of *var* gene diversity and function future work should be focused on analysis of full length sequences and the analysis of protein function and immunological responses.

5. Acknowledgements

We would like to thank the parents and guardians who participated in this study. We are grateful to Dr Bunini Manyilizu, Mrs Irene Kasiga and Sr Phillipina Hongoa of St Francis Designated Hospital for their generous support in care of the patients and sample collections. Dr Ingrid Felger for critically reviewing the manuscript. We thank Dr Kefas Mugittu, Tarsis Pius, Boniface Jullu, Krischan Bäumli for their support in laboratory analysis.

This study was supported by the Optimus Foundation and the Swiss National Science Foundation (3100 A0 – 104043/1)

5. References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology 215, 403-410.

Ariey, F., Hommel, D., Le Scanf, C., Duchemin, J.B., Peneau, C., Hulin, A., Sarthou, J.L., Reynes, J.M., Fandeur, T., Mercereau-Puijalon, O., 2001. Association of severe malaria with a specific *Plasmodium falciparum* genotype in French Guiana. J Infect Dis 184, 237-241.

Barry, A.E., Leliwa-Sytek, A., Tavul, L., Imrie, H., Migot-Nabias, F., Brown, S.M., McVean, G.A., Day, K.P., 2007. Population genomics of the immune evasion (*var*) genes of *Plasmodium falciparum*. PLoS Pathog 3, e34.

Baruch, D.I., Pasloske, B.L., Singh, H.B., Bi, X., Ma, X.C., Feldman, M., Taraschi, T.F., Howard, R.J., 1995. Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell 82, 77-87.

Breman, J.G., Holloway, C.N., 2007. Malaria surveillance counts. Am J Trop Med Hyg 77, 36-47.

Bryant, D., Moulton, V., 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21, 255-265.

Bull, P.C., Kortok, M., Kai, O., Ndungu, F., Ross, A., Lowe, B.S., Newbold, C.I., Marsh, K., 2000. *Plasmodium falciparum*-infected erythrocytes: agglutination by diverse Kenyan plasma is associated with severe disease and young host age. J Infect Dis 182, 252-259.

Bull, P.C., Marsh, K., 2002. The role of antibodies to *Plasmodium falciparum*-infected-erythrocyte surface antigens in naturally acquired immunity to malaria. Trends Microbiol 10, 55-58.

Bull, P.C., Berriman, M., Kyes, S., Quail, M.A., Hall, N., Kortok, M.M., Marsh, K., Newbold, C.I., 2005. *Plasmodium falciparum* variant surface antigen expression patterns during malaria. PLoS Pathog 1, e26.

Bustin, S.A., 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. Journal of molecular endocrinology 25, 169-193.

Chen, Q., Barragan, A., Fernandez, V., Sundstrom, A., Schlichtherle, M., Sahlen, A., Carlson, J., Datta, S., Wahlgren, M., 1998. Identification of *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite *P. falciparum*. J Exp Med 187, 15-23.

Conway, D.J., Roper, C., Oduola, A.M., Arnot, D.E., Kremsner, P.G., Grobusch, M.P., Curtis, C.F., Greenwood, B.M., 1999. High recombination rate in natural populations of *Plasmodium falciparum*. Proc Natl Acad Sci U S A 96, 4506-4511.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792-1797.

Fairhurst, R.M., Baruch, D.I., Brittain, N.J., Ostera, G.R., Wallach, J.S., Hoang, H.L., Hayton, K., Guindo, A., Makobongo, M.O., Schwartz, O.M., Tounkara, A., Doumbo, O.K., Diallo, D.A., Fujioka, H., Ho, M., Wellems, T.E., 2005. Abnormal display of PfEMP-1 on erythrocytes carrying haemoglobin C may protect against malaria. Nature 435, 1117-1121.

Falk, N., Maire, N., Sama, W., Owusu-Agyei, S., Smith, T., Beck, H.P., Felger, I., 2006. Comparison of PCR-RFLP and Genescan-based genotyping for analyzing infection dynamics of *Plasmodium falciparum*. Am J Trop Med Hyg 74, 944-950.

Felger, I., Tavul, L., Beck, H.P., 1993. *Plasmodium falciparum*: a rapid technique for genotyping the merozoite surface protein 2. Exp Parasitol 77, 372-375.

Felsenstein, J., 1985. Confidence limits on phylogenys: an approach using the bootstrap. Evolution 39, 783–791.

Fowler, E.V., Peters, J.M., Gatton, M.L., Chen, N., Cheng, Q., 2002. Genetic diversity of the DBLalpha region in *Plasmodium falciparum var* genes among Asia-Pacific isolates. Mol Biochem Parasitol 120, 117-126.

Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T.E., Scherf, A., 2000. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. Nature 407, 1018-1022.

Gamain, B., Smith, J.D., Viebig, N.K., Gysin, J., Scherf, A., 2007. Pregnancy-associated malaria: parasite binding, natural immunity and vaccine development. Int J Parasitol 37, 273-283.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.S., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M., Barrell, B., 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419, 498-511.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23, 254-267.

Kaestli, M., Cortes, A., Lagog, M., Ott, M., Beck, H.P., 2004. Longitudinal assessment of *Plasmodium falciparum var* gene transcription in naturally infected asymptomatic children in Papua New Guinea. J Infect Dis 189, 1942-1951.

Kirchgatter, K., Mosbach, R., del Portillo, H.A., 2000. *Plasmodium falciparum*: DBL-1 *var* sequence analysis in field isolates from central Brazil. Exp Parasitol 95, 154-157.

Kirchgatter, K., del Portillo, H.A., 2002. Association of severe noncerebral *Plasmodium falciparum* malaria in Brazil with expressed PfEMP1 DBL1 alpha sequences lacking cysteine residues. Mol Med 8, 16-23.

Kraemer, S.M., Gupta, L., Smith, J.D., 2003. New tools to identify *var* sequence tags and clone full-length genes using type-specific primers to Duffy binding-like domains. Mol Biochem Parasitol 129, 91-102.

Kyes, S., Taylor, H., Craig, A., Marsh, K., Newbold, C., 1997. Genomic representation of *var* gene sequences in *Plasmodium falciparum* field isolates from different geographic regions. Mol Biochem Parasitol 87, 235-238.

Kyes, S., Horrocks, P., Newbold, C., 2001. Antigenic variation at the infected red cell surface in malaria. Annu Rev Microbiol 55, 673-707.

Kyriacou, H.M., Stone, G.N., Challis, R.J., Raza, A., Lyke, K.E., Thera, M.A., Kone, A.K., Doumbo, O.K., Plowe, C.V., Rowe, J.A., 2006. Differential *var* gene transcription in *Plasmodium falciparum* isolates from patients with cerebral malaria compared to hyperparasitaemia. Mol Biochem Parasitol 150, 211-218.

Lavstsen, T., Salanti, A., Jensen, A.T., Arnot, D.E., Theander, T.G., 2003. Sub-grouping of *Plasmodium falciparum* 3D7 *var* genes based on sequence analysis of coding and non-coding regions. Malar J 2, 27.

Lavstsen, T., Magistrado, P., Hermsen, C.C., Salanti, A., Jensen, A.T., Sauerwein, R., Hviid, L., Theander, T.G., Staalsoe, T., 2005. Expression of *Plasmodium falciparum* erythrocyte membrane protein 1 in experimentally infected humans. Malar J 4, 21.

Mayor, A., Bir, N., Sawhney, R., Singh, S., Pattnaik, P., Singh, S.K., Sharma, A., Chitnis, C.E., 2005. Receptor-binding residues lie in central regions of Duffy-binding-like domains involved in red cell invasion and cytoadherence by malaria parasites. Blood 105, 2557-2563.

Miller, L.H., Baruch, D.I., Marsh, K., Doumbo, O.K., 2002. The pathogenic basis of malaria. Nature 415, 673-679.

Molyneux, M.E., Taylor, T.E., Wirima, J.J., Borgstein, A., 1989. Clinical features and prognostic indicators in paediatric cerebral malaria: a study of 131 comatose Malawian children. The Quarterly journal of medicine 71, 441-459.

Nielsen, M.A., Staalsoe, T., Kurtzhals, J.A., Goka, B.Q., Dodoo, D., Alifrangis, M., Theander, T.G., Akanmori, B.D., Hviid, L., 2002. *Plasmodium falciparum* variant surface antigen expression varies between isolates causing severe and nonsevere malaria and is modified by acquired immunity. J Immunol 168, 3444-3450.

Peters, J., Fowler, E., Gatton, M., Chen, N., Saul, A., Cheng, Q., 2002. High diversity and rapid changeover of expressed *var* genes during the acute phase of *Plasmodium falciparum* infections in human volunteers. Proc Natl Acad Sci U S A 99, 10689-10694.

Rottmann, M., Lavstsen, T., Mugasa, J.P., Kaestli, M., Jensen, A.T., Muller, D., Theander, T., Beck, H.P., 2006. Differential expression of *var* gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. Infect Immun 74, 3904-3911.

Rowe, J.A., Moulds, J.M., Newbold, C.I., Miller, L.H., 1997. *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. Nature 388, 292-295.

Rowe, J.A., Rogerson, S.J., Raza, A., Moulds, J.M., Kazatchkine, M.D., Marsh, K., Newbold, C.I., Atkinson, J.P., Miller, L.H., 2000. Mapping of the region of complement receptor (CR) 1 required for *Plasmodium falciparum* rosetting and demonstration of the importance of CR1 in rosetting in field isolates. J Immunol 165, 6341-6346.

Russo, C.A., Takezaki, N., Nei, M., 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol Biol Evol 13, 525-536.

Smith, J.D., Chitnis, C.E., Craig, A.G., Roberts, D.J., Hudson-Taylor, D.E., Peterson, D.S., Pinches, R., Newbold, C.I., Miller, L.H., 1995. Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell 82, 101-110.

Su, X.Z., Heatwole, V.M., Wertheimer, S.P., Guinet, F., Herrfeldt, J.A., Peterson, D.S., Ravetch, J.A., Wellems, T.E., 1995. The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. Cell 82, 89-100.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24, 1596-1599.

Taylor, H.M., Kyes, S.A., Newbold, C.I., 2000. *Var* gene diversity in *Plasmodium falciparum* is generated by frequent recombination events. Mol Biochem Parasitol 110, 391-397.

Trimnell, A.R., Kraemer, S.M., Mukherjee, S., Phippard, D.J., Janes, J.H., Flamoe, E., Su, X.Z., Awadalla, P., Smith, J.D., 2006. Global genetic diversity and

evolution of *var* genes associated with placental and severe childhood malaria. Mol Biochem Parasitol 148, 169-180.

Vogt, A.M., Barragan, A., Chen, Q., Kironde, F., Spillmann, D., Wahlgren, M., 2003. Heparan sulfate on endothelial cells mediates the binding of *Plasmodium falciparum*-infected erythrocytes via the DBL1alpha domain of PfEMP1. Blood 101, 2405-2411.

Voss, T.S., Thompson, J.K., Waterkeyn, J., Felger, I., Weiss, N., Cowman, A.F., Beck, H.P., 2000. Genomic distribution and functional characterisation of two distinct and conserved *Plasmodium falciparum var* gene 5' flanking sequences. Mol Biochem Parasitol 107, 103-115.

Wernersson, R., Pedersen, A.G., 2003. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res 31, 3537-3539.

WHO, 2000. Severe *falciparum* malaria. Trans R Soc Trop Med Hyg 94 Suppl 1, S1-90.

6. Tables

Table 1 Oligonucleotide primers used for amplification of different fragments of *var* genes

var Gene region	Size of amplified product	T_{anneal}	Name of primer	Primer sequence	Source
DBL1α	~500	54	DBLα-5'	5'-GCACGAAGTTTTGCAGATAT(A/T)GG-3'	
			DBLα-3'	3'-AA(A/G)TCTTC(T/G)GCCCATTCCTCGAACCA-5'	(Kaestli et al 2004)
DBL1α-CIDR	1.5 kb	52	DBLα-5'	5'-GCACGAAGTTTTGCAGATAT(A/T)GG-3'	(Kaestli et al 2004)
			CIDR1.1-3'	3'-T(C/G/T)TAGTAATTTATC(A/C/T)ATTGT-5'	
			CIDR1.2-3'	3'-T(C/G/T)TAATAAGAATTCGATTGC-5'	
upsA 5'UTR- DBL1α	1.2 kb	54	upsA-5'	5'-ATTA(C/T)ATTTGTTGTAGGTGA-3'	
			DBLα-3'	3'-AA(A/G)TCTTC(T/G)GCCCATTCCTCGAACCA-5'	
upsB 5'UTR- DBL1α	1.3 kb	52	17DBLα-5'	5'-ATGTAATTGTTGTTTTTTTTTTTGTTAGAA TATTTA AA-3'	
			DBLα-3'	3'-AA(A/G)TCTTC(T/G)GCCCATTCCTCGAACCA-5'	
psC 5'UTR- DBL1α	1.3 kb	54	5B1-5'	5'-CACATATARTACGACTAAGAAACA-3'	(Voss et al 2000)
			DBLα-3'	3'-AA(A/G)TCTTC(T/G)GCCCATTCCTCGAACCA-5'	(Kaestli et al 2004)

Table 2 Clinical and epidemiological assessment of isolates from severe and asymptomatic malaria

Isolate	Sex	Age (months)	Days between symptoms and treatment	Parasitemia (parasites/200 leukocytes)	Temperature (°C)	MUAC (cm)	PCV %	Lactate mmol/L	Glucose mmol/L	Blantyre score
Severe										
ISM2	M	40	3	3120	38.6	17	27	2.5	2.3	2
ISM3	M	36	2	2574	39.7	17	16	2.4	5.9	3
ISM11	M	33	2	7344	37.0	17	30	4.5	3.7	3
ISM16	M	16	3	2484	38.6	14	26	2.1	2.7	3
IMS33	F	24	3	1316	37.4	16	22	3.1	3.0	2
ISM48	M	36	2	4713	38.9	17	31	3.0	5.0	3
ISM49	M	14	4	1907	39.9	16	23	1.8	6.6	3
ISM51	F	16	4	9999	40.0	16	21	5.0	8.4	2
Asymptomatic										
IAM5	M	24	NA	830	36.7	ND	20	3.4	5.8	NA
IAM7	F	33	NA	704	36.6	ND	14	3.3	3.9	NA
IAM10	M	59	NA	70	37.5	ND	22	2.4	4.2	NA
IAM11	F	56	NA	68	36.7	ND	23	4.6	4.4	NA
IAM12	F	47	NA	690	36.4	ND	26	4.5	8.0	NA
IAM17	M	59	NA	360	37.3	ND	28	2.5	4.9	NA
IAM18	F	52	NA	360	37.3	ND	18	3.1	4.3	NA

A – Not applicable, ND – Not determined

Table 3a Summary of analysed sequences of different transcribed *var* DBL1α sequences

Isolates	RT-PCR, Cloned in pGEMT Vector, PCR screened 96 clones picked for sequencing	Sequences generated per isolate	Number of distinct DBLα var per isolate	Predominant gene blasted vs 3D7	Group holomogy to 3D7	Bulls' signature and group	MOI
Severe							
ISM 2	48	42	23	PF08_0141	A	LFLG-IREY-KAIT-2-LTNL	2
ISM 3	48	41	22	PFF0010w	B/A	LYLD-FREY-KAIT-2-PTNL	3
ISM 11	60	50	20	PFD1005c	B/C	LFIG-LRED-KALT-4-PTYF	2
ISM 16	48	36	18	PFD0020c	A	MFKR-LRED-RAIT-2-PTNL	1
ISM 33	60	50	15	PFF0010w	A	LFLG-VREY-KAIT-2-LTNL	3
ISM 48	48	47	12	PF08_0141	A	MFLG-IREY-KALT-2-PTNL	3
ISM 49	48	44	17	PFL1830c	B	LYLG-LRED-KAIT-4-PTYF	3
Asymptomatic							
IAM 5	48	46	25	PF08_0141	A	MFLG-IREY-KALT-2-PTNL	3
IAM 7	48	45	8	PFL 1830c	B	LYLG-LRED-KALT-4-PTYF	2
IAM 10	48	38	23	PFD0615c	C	LFIG-LRED-EAIT-4-PTNF	3
IAM 11	48	43	15	PFL2665c	B	LYRG-LRED-NAII-3-LTNF	3
IAM 12	48	45	20	PFL1955w	B/C	LYLG-LRED-KAIT-4-PTYF	2
IAM 17	48	46	24	PFA0005w	B	LYLG-LRED-EAIT-4-PTYF	2
IAM 18	48	42	16	PFA0005w	B	LYLG-LRED-KAIT-4-PTYF	2
	Total	665					

MOI- Multiplicity of infections

Table 3b Summary of analysed upstream sequences transcribed by different isolates

Domain	Isolate	RT-PCR, Cloned in pCR®4-TOPO vector, PCR screened 96 clones picked for sequencing	Sequences generated	Number of distinct sequences per isolate	Predominant *var* DBL blasted vs 3D7	Holomogy to 3D7	Whole fragment blasted vs 3D7	Homology to 3D7
5B1	ISM 2	48	19	4	PFD0625	C	PFD0625	C
	ISM 33	48	26	8	MAL7P1.56	C	MAL7P1.56	C
17DBL	IAM 18	48	39	10	PFA0005w	B	PFA0005w	B
	ISM 49	48	27	6	PFL0005w	B	PFL0020w	B/A
	ISM 3	48	29	3	PFD10005c	B/C	PFD0020c	A
DBL-CIDR	ISM 11	48	35	5	PF08_0141	A	PFD0020c	A
	ISM 51	48	39	13	PFL0020w	B/A	PFD0020c	A
	IAM 17	48	31	4	PFE1640w	*var* 1	PFE1640w	*var* 1
upsA	ISM 2	48	18	6	PFD0020c	A	PFD0020c	A
	ISM 3	60	44	5	PF08_0141	A	PFD1235w	A
	ISM 11	48	35	5	PFD0020c	A	PFD0020c	A
	ISM 16	60	45	3	PFD0020c	A	PFD0020c	A
	ISM 49	60	48	3	PFL1820w	A	PFE1640w	*var* 1
	ISM 51	48	40	7	PFD0020c	A	PFD0020c	A
		Total	665					

7. Figure Legends

Figure. 1 Distribution of unique sequence types (STs) of DBL1α in clinical isolates

Blue columns represent STs found in multiple samples from both AM and SM groups. Red columns represent STs found in multiple samples within the SM group. Pink columns represent STs found in multiple samples within the AM group. Yellow columns represent STs specific to each individual isolate

Figure. 2 Distribution of DBL-1α sequences into cys/PoLV groups by clinical status.

SM DBL-1α sequences had more than 50% Cys2 sequence tags (1-3 groups) compared to 27 % in AM isolates

Figure. 3 Distribution of PoLV motifs within clinical isolates SM (red bars) and in AM (white bars)

Figure. 4 Cumulative diversity curves for DBL-1α sequences from Ifakara.

The cumulative curve for DBL-1α was determined by simulation of the number of unique sequences as a function of the number of patient samples as a function of the number of patient samples. For each number of patient samples the statistics value were obtained from simulations of all possible sample combinations

Figure. 5 Phylogenetic network showing the comparison of 3 dominant DBL-1α sequence tags transcribed from clinical isolates, generated using Neighbour-Net (Bryant & Moulton 2004). Sequences transcribed by isolates with severe malaria (ISM, blue) and asymptomatic malaria (IAM, red) are compared. The sequences fall into two major clades separated with dotted line, the upper cluster formed 2 subgroups of sequences isolated from severe patients, one with group A and the remaining *var* group B/A homology to 3D7. Isolates from asymptomatic patient majority of the sequences clustering together and were homology to group B, B/C and C of 3D7 genome

Figure. 6 Phylogenetic comparison of DBLα from 3D7 genome and 3 dominant DBLα sequence tags transcribed from each clinical isolates. A neighbor-joining tree was generated based amplified DBL-1α fragments. The bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985), using pairwise deletion of amino acid sequences with p-distance. Phylogenetic analyses were conducted in MEGA4 (Tamura et al 2007). Sequences transcribed by isolates from children with severe malaria (ISM, pink), asymptomatic malaria (IAM, black) and 3D7 genes (group A, red; group B, blue; group C, green.

Figure. 7 Phylogenetic comparison of *var* group A from 3D7 genome and 3 dominant *var* group A amplified from clinical isolates. A neighbor-joining tree was generated based upon the predicted protein start site in the N-terminal segment (NTS) domain to the first DBL-1α H block. The bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985), using pairwise deletion of amino acid sequences with p-distance. Phylogenetic analyses were conducted in MEGA4 (Tamura et al 2007). Sequences transcribed by isolates from children with severe malaria (ISM, blue), asymptomatic

Figure 1

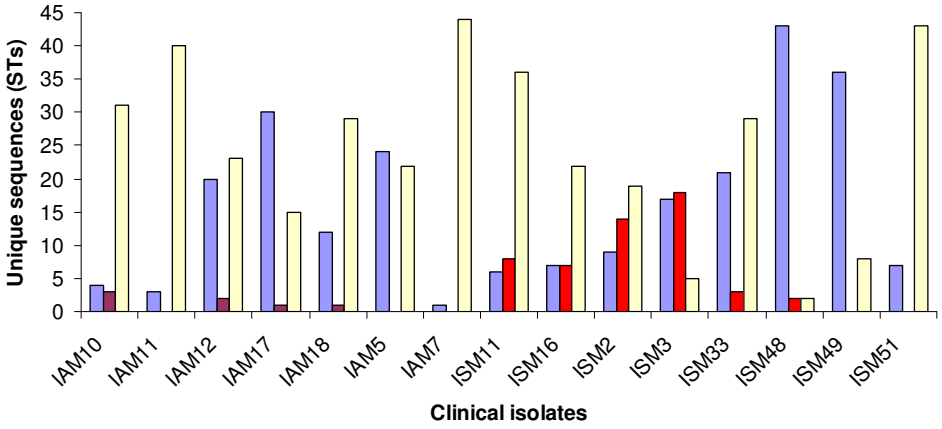


Figure 2

.

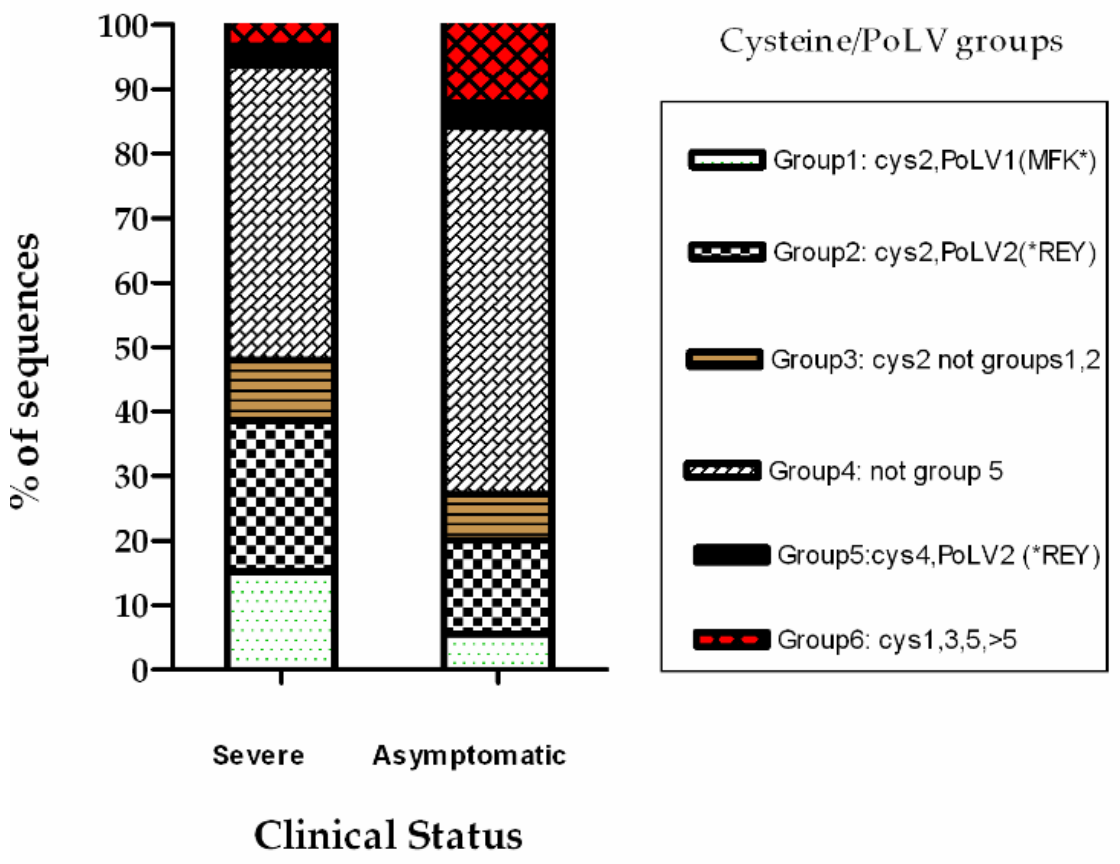


Figure 3

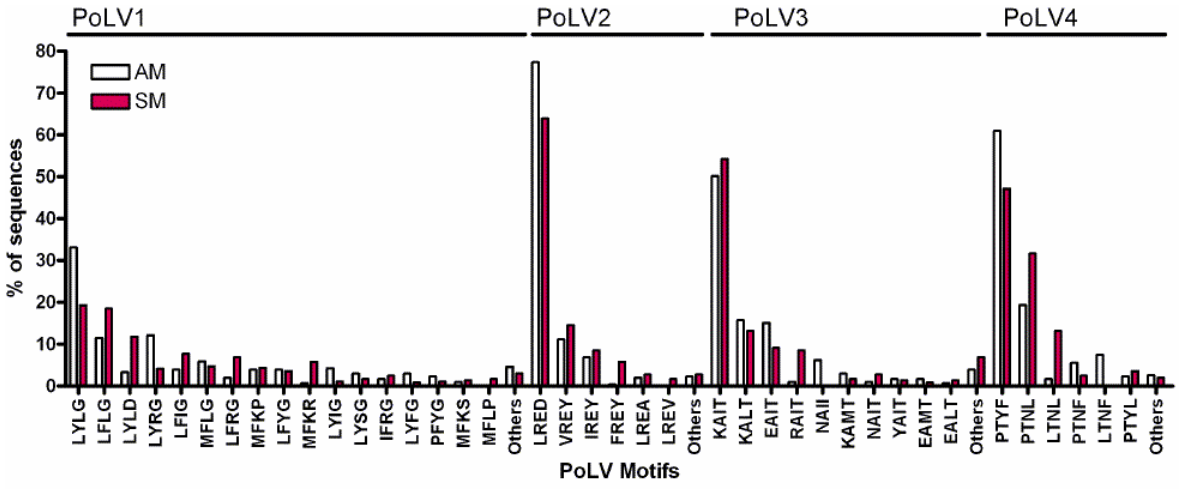


Figure 4

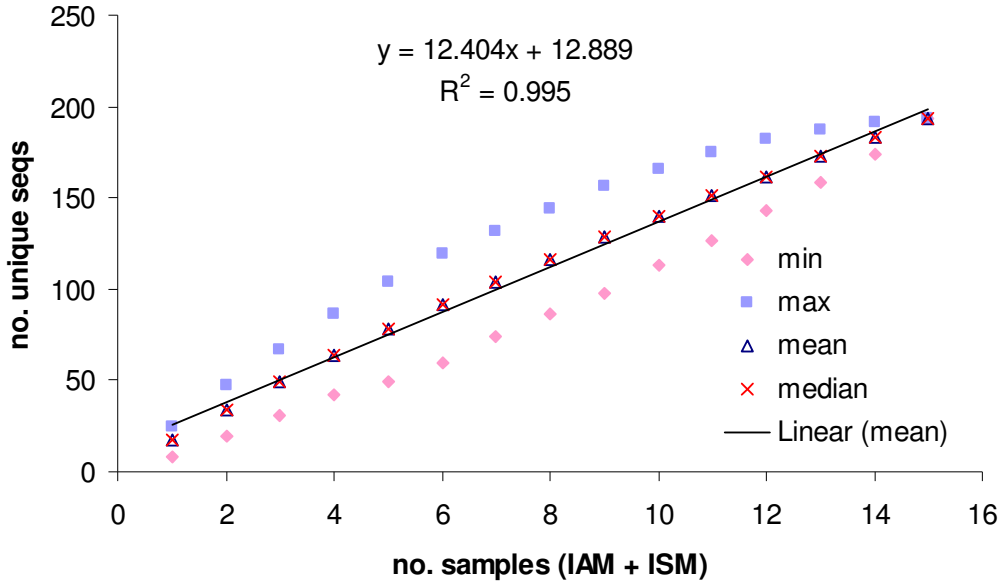


Figure 5

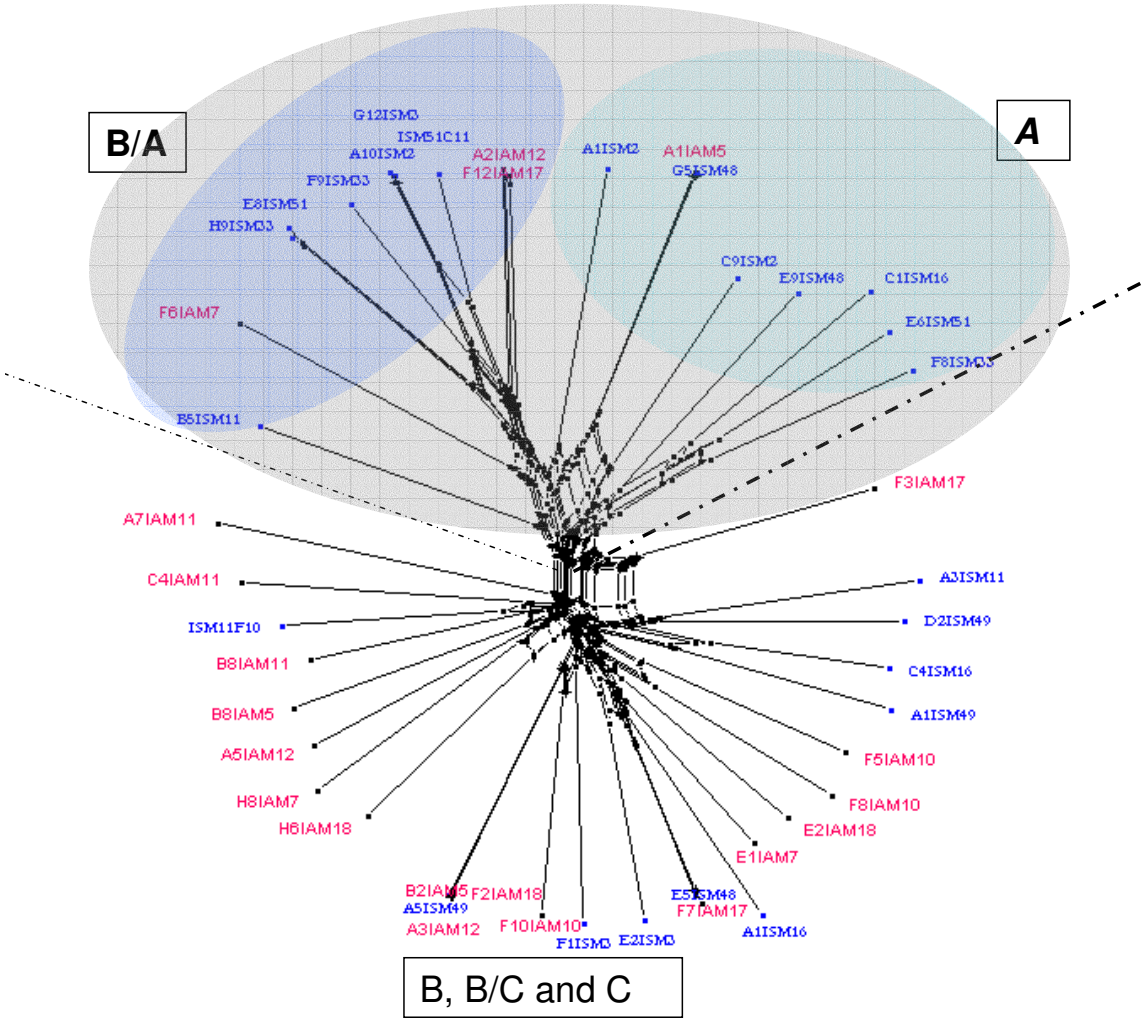

B/A
A
G12ISM3
ISM51C11
A10ISM2
A2IAM12
F12IAM17
A1ISM2
A1IAM5
G5ISM48
F9ISM33
E8ISM51
H9ISM33
C9ISM2
E9ISM48
C1ISM16
E6ISM51
F8ISM33
F6IAM7
B5ISM11
F3IAM17
A7IAM11
C4IAM11
A3ISM11
ISM11F10
D2ISM49
B8IAM11
C4ISM16
B8IAM5
A1ISM49
A5IAM12
F5IAM10
H8IAM7
F8IAM10
H6IAM18
E2IAM18
E1IAM7
B2IAM5
F2IAM18
A5ISM49
E5ISM48
F7IAM17
A3IAM12
F10IAM10
A1ISM16
F1ISM3
E2ISM3
B, B/C and C

Figure 6

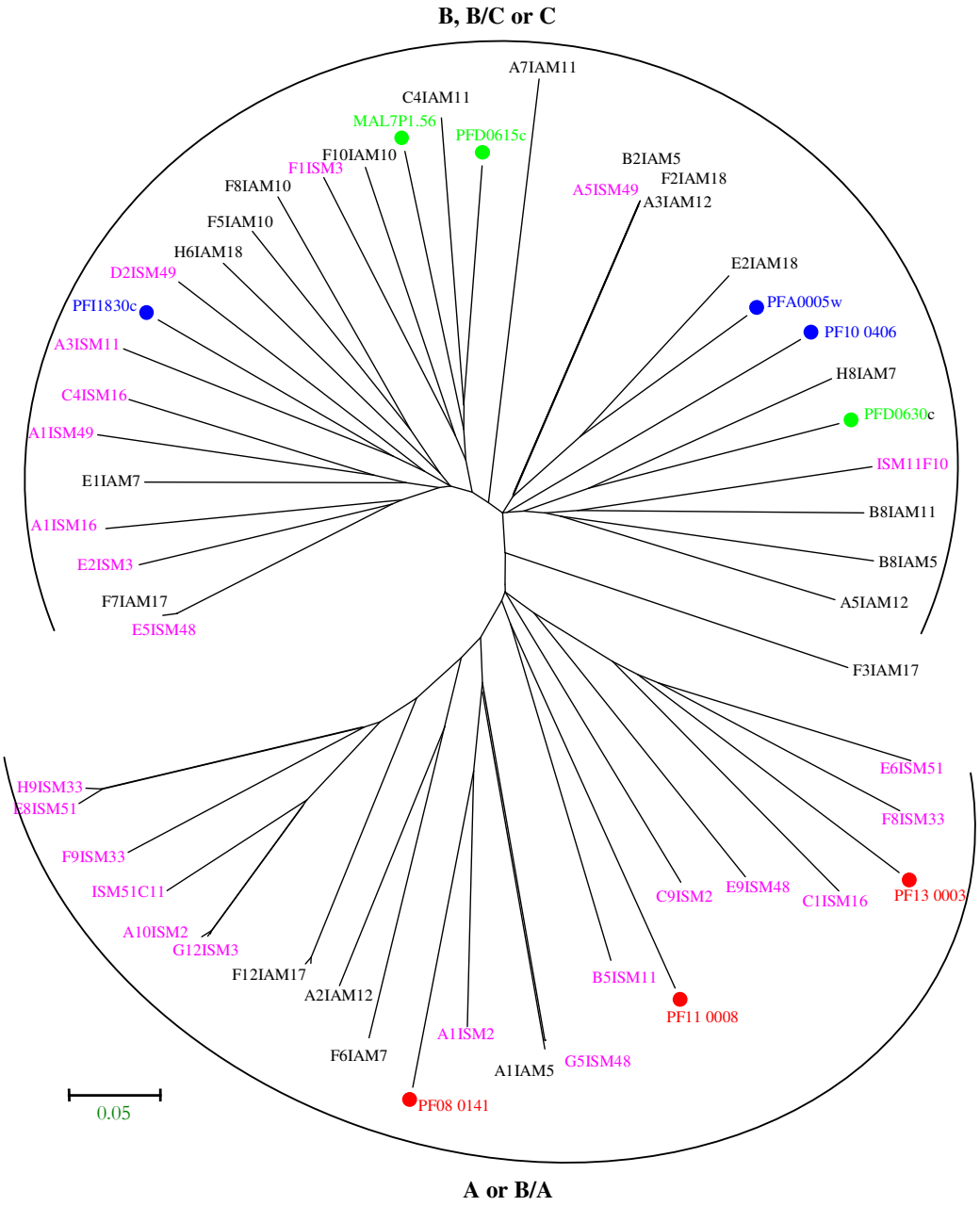
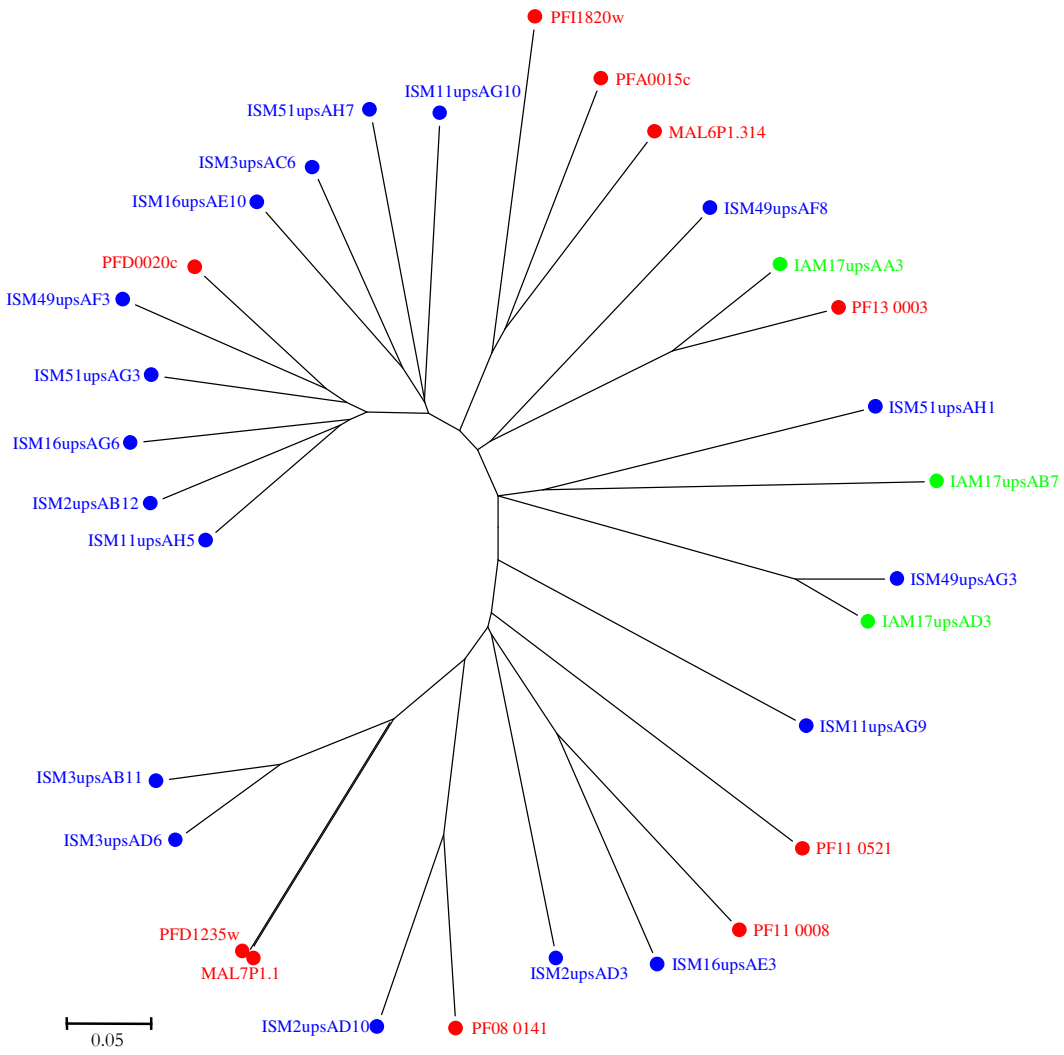

B, B/C or C
A7IAM11
C4IAM11
MAL7P1.56
PFD0615c
F10IAM10
F1ISM3
F8IAM10
F5IAM10
H6IAM18
D2ISM49
PFI1830c
A3ISM11
C4ISM16
A1ISM49
E1IAM7
A1ISM16
E2ISM3
F7IAM17
E5ISM48
B2IAM5
F2IAM18
A5ISM49
A3IAM12
E2IAM18
PFA0005w
PF10 0406
H8IAM7
PFD0630c
ISM11F10
B8IAM11
B8IAM5
A5IAM12
F3IAM17
E6ISM51
F8ISM33
PF13 0003
C1ISM16
E9ISM48
C9ISM2
B5ISM11
PF11 0008
G5ISM48
A1IAM5
A1ISM2
PF08 0141
F6IAM7
A2IAM12
F12IAM17
G12ISM3
A10ISM2
ISM51C11
F9ISM33
E8ISM51
H9ISM33
0.05
A or B/A

Figure 7


PFI1820w
PFA0015c
ISM11upsAG10
ISM51upsAH7
MAL6P1.314
ISM3upsAC6
ISM16upsAE10
ISM49upsAF8
PFD0020c
IAM17upsAA3
ISM49upsAF3
PF13 0003
ISM51upsAG3
ISM51upsAH1
ISM16upsAG6
IAM17upsAB7
ISM2upsAB12
ISM11upsAH5
ISM49upsAG3
IAM17upsAD3
ISM11upsAG9
ISM3upsAB11
ISM3upsAD6
PF11 0521
PFD1235w
PF11 0008
MAL7P1.1
ISM2upsAD3
ISM16upsAE3
ISM2upsAD10
PF08 0141
0.05

Chapter Five

General Discussion and Conclusion

Chapter Five

5.0 General Discussion and Conclusion

Plasmodium falciparum is unique among parasites that cause malaria in humans due to its ability to induce cytoadhesion of infected red blood cells. This property is a consequent of the exposure of PfEMP1 protein on the surfaces of infected cells. The resulting adhesive properties are believed to be the major virulence determinants of *P. falciparum* infections. PfEMP1 plays a large role in host-parasite interactions.

Despite years of research, very little is known about changes that occur in the host-parasite relationship as naturally acquired antimalarial immunity develops. Molecular tools for measuring changes in the parasite as it adapts to the development of clinical immunity *in vivo* are still lacking. Such a tool would provide a powerful means of dissecting the protective components of host response, a first step in the identification of new vaccine candidates. One of the limitations of studying host-parasite interaction has been possible to overcome by using parasite isolates from peripheral blood, since there is no equivalent or useful animal model system for mimicking cytoadherence in the host. It is possible to argue though that parasites isolated from peripheral blood might not reflect what is happening in deep tissue and organs of the host.

Both clinical and sero-epidemiological studies have shown that parasites from patients with severe malaria express a different subset of surface antigen that are more frequently recognized by sera from malaria exposed individuals, including young children, than parasite antigens from older children with mild malaria (Bull et al 2000, Nielsen et al 2002). It has been shown that this subset of surface antigens is serologically conserved among different geographical regions (Nielsen et al 2004). Therefore, it is critical to identify the molecular phenotypes and the genetic diversity of such a subset in-order to develop a disease ameliorating vaccine or for other therapeutic interventions.

Chapter Five

In this research, two studies on *var* gene expression during malaria infection are described. In the first study, the differential expression levels of PfEMP1 encoded by *var* gene group A, B, or C were quantified. The real-time quantitative PCR procedure was used to compare the distribution of *var* gene transcripts of *var* group A, B and C among children with asymptomatic malaria (AM), uncomplicated malaria (UM) and severe malaria (SM). In the second study, the case definition of severe malaria was restricted by using children with cerebral malaria (Blantyre coma score ≤ 3). The genetic diversity of expressed *var* genes from children with severe malaria was compared with that of children with asymptomatic malaria infection.

Both studies were carried out in Ifakara, a semi-rural area, in Kilombero valley, in Southern Tanzania. Ifakara is an area of moderate perennial *P. falciparum* transmission surrounded by areas of more intense transmission. Malaria is known to be the leading cause of morbidity and mortality in children under age of 5 years (Schellenberg et al 2004).

5.1 Differential Expression of *var* gene Groups

Transcripts of *var* group A and B genes were more abundant in patients with severe malaria than in those with mild malaria. It was clearly demonstrated that the transcript abundances of *var* group A and B gene were higher for children with clinical malaria than for children with asymptomatic infections. *var* group C had no association with any clinical manifestation.

Recent studies on *var* gene transcription and malaria severity in clinical isolates have yielded conflicting results. Differences in epidemiology, *var* classification, and severe disease characterization make comparison across studies difficult.

Chapter Five

Severe malaria encompasses a variety of clinical syndromes (*i.e* cerebral malaria, severe anaemia, respiratory distress and prostration). These syndromes may have different underlying cytoadherent properties of parasites and different pathogenic mechanisms. Significant associations between *var* gene groups and clinical disease may be masked unless strictly defined clinical groups are assessed along with suitable control groups (Kurtzhals et al 2001). This represented one of the major challenges in the first study as we considered severe malaria as a single entity to facilitate the analysis. Large sample size and strict clinical definition of severe malaria should provide strong evidence of a particular *var* gene group expression with a particular subset of severe malaria syndrome. In the study, peripheral blood samples were used to examine *P. falciparum var* gene transcription profiles in clinical isolates. There is a big debate whether parasites circulating in peripheral blood adequately represent the sequestered parasites, and the disease associated phenotype. Ideally, studies aimed at examining the relationship between *var* gene transcription and a particular disease syndrome would preferably examine sequestered parasites. However, these are only accessible in postmortem samples, and conducting such studies are technically and ethically challenging. There is controversial data on phenotypes of circulating parasites and adhering parasites from human placental studies. Two studies reported identical binding phenotype in the placental and in the peripheral circulation (Ofori et al 2003, Tuikue Ndam et al 2004), whereas other studies showed antigenically distinct parasites (Beeson et al 1999, Fried & Duffy 1996). Three recent studies from Malawi (Dembo et al 2006, Dobano et al 2007, Montgomery et al 2006) indicates that sequestered parasites are usually similar to those in peripheral circulation, and that parasite genotypes in cerebral malaria patients are homogenously distributed throughout the body. In another study, (Montgomery et al 2007) by using pediatric postmortem samples showed that the *var* genes expressed in the brain, lungs and heart are subsets of *var* transcripts found in the spleen, which might represent peripheral circulation.

Yet, the exact proportion of sequestered parasites present in peripheral circulation remains uncertain. Thus, although we hypothesize that severe malaria is maintained by a homogenous parasite population expressing multi-adhesive variant surface antigens (VSA), or parasites expressing a VSA subset, which mediates sequestration independent of endothelial receptors it might be quite challenging to identify these.

5.2 Genetic Diversity of Expressed *Plasmodium falciparum var* genes

This study showed a wide genetic diversity of the *var* gene with minimal overlaps between isolates. DBL-1α sequences from SM had a reduced number of cysteines residues, and most belonged to *var* gene group A and B/A with homology to 3D7. Detailed analysis of the sequences showed all isolates had a predominant sequence as well as minor transcripts and unique sequence types. Although, the dominant transcript differed between isolates, it is uncertain whether the dominance of particular *var* genes within a patient truly reflected the differential expression levels, or whether it is due to biased amplification, cloning or unequal distribution of circulating stages. For either of these assumptions to be true, we would expect that the same bias would be observed in all patient isolates and therefore we would not see a different dominant sequence in each isolate. The differences in *var* transcripts in each isolate suggest that each parasite genome has its own form of *var* gene repertoires. This has an implication for acquiring long-term protective immunity. Infection with a particular strain of *P. falciparum* does not necessarily give protection to subsequent infections. The DBL-1α and upstream sequences variability within an isolate were similar to that found between different isolates regardless of the group (AS or SM). This suggests that the level of diversity detected is representative of all the possible sequence diversity existing within each isolate. Using the collector's curve analysis on all

DBL-1α sequences generated, the total repertoire of *var* gene sampled from a local population of Ifakara area was estimated. This is a region of circa 50 km in diameter surrounding the St Francis Designated District Hospital, where the study subjects were recruited. The analysis showed that the *var* gene repertoire in the sampled population is unlimited as the saturation point could not be reached. This result indicates a vast amount of *var* gene diversity in the Ifakara area, and is evidence for the vast archive of antigenic diversity in *P. falciparum* and this might explain why immunity to malaria is non-sterilizing and develops slowly. Whether this is the representation of the *var* gene diversity of Ifakara or Tanzania in general still remains to be elucidated. Whether increased sample size and inclusion of more than only one *var* gene domain would have helped to elucidate the total *var* gene diversity is a challenge to be solved. However, the *var* genes were found to have an unlimited repertoire in this study, and further cluster analysis revealed that AM isolates were more diverse with more singletons compared to SM isolates. This shows that the diversity of *var* genes within SM patients was more restricted, hence could explain why immunity to severe childhood malaria develops earlier after a few episodes.

5.3 Conclusions

The pathogenesis of *falciparum* malaria involves complex interactions of host and parasite factors, further complicated by the fact that antigenic and adhesive properties of circulating parasites may be quite different to those sequestered. The increased body of evidence has demonstrated that circulating parasites represent the sequestered parasites. Despite the importance of *var* genes in malaria pathogenesis, most of the expression studies have been done either on *in vitro* cultured field isolates or on reference strains. In *in vitro* cultures, the parasites are grown without the selective pressure of the host's immune system, therefore no changes of their intrinsic phenotype occurs, consequently becoming homogeneous with respect to the PfEMP1 variant expressed (Frank et al 2007, Peters et al 2007). Scherf et al observed subtelomeric deletion in cultured isolates which resulted into loss of functional genes. (Scherf et al 1992). Our studies on *var* gene transcriptions on naturally infected children were performed without culturing the parasite isolates, therefore, providing a good insight of the complex nature of this family of genes.

The diversity of *var* genes is massive and encourages broader analysis of *var* subpopulations from other geographical areas with different malaria epidemiology. It will be important to define the total *var* gene repertoire in a given geographical area or globally, to facilitate the analysis of temporal changes over time or as new *P. falciparum* strains invade endemic areas. It might seem unreasonable to use the *var* genes family as the basis of vaccine or therapeutic targets owing to its extraordinarily high polymorphism and complex nature. However, *var2csa* is a promising vaccine candidate for pregnancy-associated malaria (Avril et al 2006, Rogerson et al 2007). With better understanding of *var* gene diversity and function, we might be able to take advantage of specific *var*

subtypes associated with severe malaria, since the *var* gene family remains the best defined factor contributing to malaria pathogenesis. The completion of human and *P. falciparum* genomes, high throughput methodologies and reverse genetics, along with clinical data from different epidemiological settings will hopefully lead to a better understanding of the role of these complex interactions in different clinical syndromes. Most studies of *var* genes have focused on small fragments of DBL1α domain. With the high throughput technologies for sequencing available (Solexa and 454 life science), broader analysis of full-length *var* genes, the analysis of protein function and immunological responses will provide invaluable information toward new opportunities for interventions to treat or prevent severe malaria.

6.0 References

References of Chapter One, Two and Five

Aguiar JC, Albrecht GR, Cegielski P, Greenwood BM, Jensen JB, et al. 1992. Agglutination of *Plasmodium falciparum*-infected erythrocytes from east and west African isolates by human sera from distant geographic regions. *Am J Trop Med Hyg* 47: 621-32

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215: 403-10

Ariey F, Hommel D, Le Scanf C, Duchemin JB, Peneau C, et al. 2001. Association of severe malaria with a specific *Plasmodium falciparum* genotype in French Guiana. *J Infect Dis* 184: 237-41

Avril M, Gamain B, Lepolard C, Viaud N, Scherf A, Gysin J. 2006. Characterization of anti-var2CSA-PfEMP1 cytoadhesion inhibitory mouse monoclonal antibodies. *Microbes Infect* 8: 2863-71

Barnwell JW, Ockenhouse CF, Knowles DM, 2nd. 1985. Monoclonal antibody OKM5 inhibits the in vitro binding of *Plasmodium falciparum*-infected erythrocytes to monocytes, endothelial, and C32 melanoma cells. *J Immunol* 135: 3494-7

Barragan A, Kremsner PG, Wahlgren M, Carlson J. 2000. Blood group A antigen is a coreceptor in *Plasmodium falciparum* rosetting. *Infect Immun* 68: 2971-5

Barry AE, Leliwa-Sytek A, Tavul L, Imrie H, Migot-Nabias F, et al. 2007. Population genomics of the immune evasion (*var*) genes of *Plasmodium falciparum*. *PLoS Pathog* 3: e34

Baruch DI, Pasloske BL, Singh HB, Bi X, Ma XC, et al. 1995. Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* 82: 77-87

Beeson JG, Brown GV, Molyneux ME, Mhango C, Dzinjalamala F, Rogerson SJ. 1999. *Plasmodium falciparum* isolates from infected pregnant women and children are associated with distinct adhesive and antigenic properties. *J Infect Dis* 180: 464-72

Beeson JG, Rogerson SJ, Cooke BM, Reeder JC, Chai W, et al. 2000. Adhesion of *Plasmodium falciparum*-infected erythrocytes to hyaluronic acid in placental malaria. *Nat Med* 6: 86-90

Berendt AR, Simmons DL, Tansey J, Newbold CI, Marsh K. 1989. Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for *Plasmodium falciparum*. *Nature* 341: 57-9

Breman JG. 2001. The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am J Trop Med Hyg* 64: 1-11

References

Breman JG, Egan A, Keusch GT. 2001. The intolerable burden of malaria: a new look at the numbers. *Am J Trop Med Hyg* 64: iv-vii

Breman JG, Holloway CN. 2007. Malaria surveillance counts. *Am J Trop Med Hyg* 77: 36-47

Bryant D, Moulton V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* 21: 255-65

Buffet PA, Gamain B, Scheidig C, Baruch D, Smith JD, et al. 1999. *Plasmodium falciparum* domain mediating adhesion to chondroitin sulfate A: a receptor for human placental infection. *Proc Natl Acad Sci U S A* 96: 12743-8

Bull PC, Berriman M, Kyes S, Quail MA, Hall N, et al. 2005. *Plasmodium falciparum* variant surface antigen expression patterns during malaria. *PLoS Pathog* 1: e26

Bull PC, Kortok M, Kai O, Ndungu F, Ross A, et al. 2000. *Plasmodium falciparum*-infected erythrocytes: agglutination by diverse Kenyan plasma is associated with severe disease and young host age. *J Infect Dis* 182: 252-9

Bull PC, Lowe BS, Kortok M, Marsh K. 1999. Antibody recognition of *Plasmodium falciparum* erythrocyte surface antigens in Kenya: evidence for rare and prevalent variants. *Infect Immun* 67: 733-9

Bull PC, Marsh K. 2002. The role of antibodies to *Plasmodium falciparum*-infected-erythrocyte surface antigens in naturally acquired immunity to malaria. *Trends Microbiol* 10: 55-8

Bustin SA. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 25: 169-93

Carlson J, Wahlgren M. 1992. *Plasmodium falciparum* erythrocyte rosetting is mediated by promiscuous lectin-like interactions. *J Exp Med* 176: 1311-7

Chakravorty SJ, Craig A. 2005. The role of ICAM-1 in *Plasmodium falciparum* cytoadherence. *Eur J Cell Biol* 84: 15-27

Chen Q, Barragan A, Fernandez V, Sundstrom A, Schlichtherle M, et al. 1998. Identification of *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite *P. falciparum*. *J Exp Med* 187: 15-23

Chen Q, Schlichtherle M, Wahlgren M. 2000. Molecular aspects of severe malaria. *Clin Microbiol Rev* 13: 439-50

Chiwakata CB, Manegold C, Bonicke L, Waase I, Julch C, Dietrich M. 2001. Procalcitonin as a parameter of disease severity and risk of mortality in patients with *Plasmodium falciparum* malaria. *J Infect Dis* 183: 1161-4

Clough B, Atilola FA, Pasvoi G. 1998. The role of rosetting in the multiplication of *Plasmodium falciparum*: rosette formation neither enhances nor targets parasite invasion into uninfected red cells. *Br J Haematol* 100: 99-104

Collins WE, Jeffery GM. 2005. *Plasmodium ovale*: parasite and disease. *Clin Microbiol Rev* 18: 570-81

References

Conway DJ, Roper C, Oduola AM, Arnot DE, Kremsner PG, et al. 1999. High recombination rate in natural populations of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* 96: 4506-11

Cox-Singh J, Davis TM, Lee KS, Shamsul SS, Matusop A, et al. 2008. *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clin Infect Dis* 46: 165-71

Craig AG, Pinches R, Khan S, Roberts DJ, Turner GD, et al. 1997. Failure to block adhesion of *Plasmodium falciparum*-infected erythrocytes to ICAM-1 with soluble ICAM-1. *Infect Immun* 65: 4580-5

Degen R, Weiss N, Beck HP. 2000. *Plasmodium falciparum*: cloned and expressed CIDR domains of PfEMP1 bind to chondroitin sulfate A. *Exp Parasitol* 95: 113-21

Dembo EG, Phiri HT, Montgomery J, Molyneux ME, Rogerson SJ. 2006. Are *Plasmodium falciparum* parasites present in peripheral blood genetically the same as those sequestered in the tissues? *Am J Trop Med Hyg* 74: 730-2

Dobano C, Rogerson SJ, Taylor TE, McBride JS, Molyneux ME. 2007. Expression of merozoite surface protein markers by *Plasmodium falciparum*-infected erythrocytes in peripheral blood and tissues of children with fatal malaria. *Infect Immun* 75: 643-52

Duarte MI, Corbett CE, Boulos M, Amato Neto V. 1985. Ultrastructure of the lung in *falciparum* malaria. *Am J Trop Med Hyg* 34: 31-5

Duffy MF, Maier AG, Byrne TJ, Marty AJ, Elliott SR, et al. 2006. VAR2CSA is the principal ligand for chondroitin sulfate A in two allogeneic isolates of *Plasmodium falciparum*. *Mol Biochem Parasitol* 148: 117-24

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792-7

Fairhurst RM, Baruch DI, Brittain NJ, Ostera GR, Wallach JS, et al. 2005. Abnormal display of PfEMP-1 on erythrocytes carrying haemoglobin C may protect against malaria. *Nature* 435: 1117-21

Falk N, Maire N, Sama W, Owusu-Agyei S, Smith T, et al. 2006. Comparison of PCR-RFLP and Genescan-based genotyping for analyzing infection dynamics of *Plasmodium falciparum*. *Am J Trop Med Hyg* 74: 944-50

Felger I, Tavul L, Beck HP. 1993. *Plasmodium falciparum*: a rapid technique for genotyping the merozoite surface protein 2. *Exp Parasitol* 77: 372-5

Felsenstein J. 1985. Confidence limits on phylogenys: an approach using the bootstrap. *Evolution* 39: 783–91

Fernandez V, Treutiger CJ, Nash GB, Wahlgren M. 1998. Multiple adhesive phenotypes linked to rosetting binding of erythrocytes in *Plasmodium falciparum* malaria. *Infect Immun* 66: 2969-75

Fernandez V, Wahlgren M. 2002. Rosetting and autoagglutination in *Plasmodium falciparum*. *Chem Immunol* 80: 163-87

Fidock DA, Rosenthal PJ, Croft SL, Brun R, Nwaka S. 2004. Antimalarial drug discovery: efficacy models for compound screening. *Nat Rev Drug Discov* 3: 509-20

Fowler EV, Peters JM, Gatton ML, Chen N, Cheng Q. 2002. Genetic diversity of the DBLalpha region in *Plasmodium falciparum var* genes among Asia-Pacific isolates. *Mol Biochem Parasitol* 120: 117-26

Frank M, Dzikowski R, Amulic B, Deitsch K. 2007. Variable switching rates of malaria virulence genes are associated with chromosomal position. *Mol Microbiol* 64: 1486-98

Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, et al. 2000. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* 407: 1018-22

Frevert U, Sinnis P, Cerami C, Shreffler W, Takacs B, Nussenzweig V. 1993. Malaria circumsporozoite protein binds to heparan sulfate proteoglycans associated with the surface membrane of hepatocytes. *J Exp Med* 177: 1287-98

Fried M, Duffy PE. 1996. Adherence of *Plasmodium falciparum* to chondroitin sulfate A in the human placenta. *Science* 272: 1502-4

Gamain B, Smith JD, Viebig NK, Gysin J, Scherf A. 2007. Pregnancy-associated malaria: parasite binding, natural immunity and vaccine development. *Int J Parasitol* 37: 273-83

Gamain B, Trimnell AR, Scheidig C, Scherf A, Miller LH, Smith JD. 2005. Identification of multiple chondroitin sulfate A (CSA)-binding domains in the *var2CSA* gene transcribed in CSA-binding parasites. *J Infect Dis* 191: 1010-3

Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419: 498-511

Gaur D, Mayer DC, Miller LH. 2004. Parasite ligand-host receptor interactions during invasion of erythrocytes by *Plasmodium* merozoites. *Int J Parasitol* 34: 1413-29

Gupta S, Snow RW, Donnelly CA, Marsh K, Newbold C. 1999. Immunity to non-cerebral severe malaria is acquired after one or two infections. *Nat Med* 5: 340-3

Heddini A, Pettersson F, Kai O, Shafi J, Obiero J, et al. 2001. Fresh isolates from children with severe *Plasmodium falciparum* malaria bind to multiple receptors. *Infect Immun* 69: 5849-56

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23: 254-67

Jensen AT, Magistrado P, Sharp S, Joergensen L, Lavstsen T, et al. 2004. *Plasmodium falciparum* associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A *var* genes. *J Exp Med* 199: 1179-90

References

Kaestli M, Cockburn IA, Cortes A, Baea K, Rowe JA, Beck HP. 2006. Virulence of malaria is associated with differential expression of *Plasmodium falciparum var* gene subgroups in a case-control study. *J Infect Dis* 193: 1567-74

Kaestli M, Cortes A, Lagog M, Ott M, Beck HP. 2004. Longitudinal assessment of *Plasmodium falciparum var* gene transcription in naturally infected asymptomatic children in Papua New Guinea. *J Infect Dis* 189: 1942-51

Kappe SH, Buscaglia CA, Nussenzweig V. 2004. *Plasmodium* sporozoite molecular cell biology. *Annu Rev Cell Dev Biol* 20: 29-59

Kirchgatter K, del Portillo HA. 2002. Association of severe noncerebral *Plasmodium falciparum* malaria in Brazil with expressed PfEMP1 DBL1 alpha sequences lacking cysteine residues. *Mol Med* 8: 16-23

Kirchgatter K, Mosbach R, del Portillo HA. 2000. *Plasmodium falciparum*: DBL-1 *var* sequence analysis in field isolates from central Brazil. *Exp Parasitol* 95: 154-7

Kochar DK, Pakalapati D, Kochar SK, Sirohi P, Khatri MP, et al. 2007. An unexpected cause of fever and seizures. *Lancet* 370: 908

Kraemer SM, Gupta L, Smith JD. 2003. New tools to identify *var* sequence tags and clone full-length genes using type-specific primers to Duffy binding-like domains. *Mol Biochem Parasitol* 129: 91-102

Kraemer SM, Kyes SA, Aggarwal G, Springer AL, Nelson SO, et al. 2007. Patterns of gene recombination shape *var* gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. *BMC Genomics* 8: 45

Kraemer SM, Smith JD. 2003. Evidence for the importance of genetic structuring to the structural and functional specialization of the *Plasmodium falciparum var* gene family. *Mol Microbiol* 50: 1527-38

Kraemer SM, Smith JD. 2006. A family affair: *var* genes, PfEMP1 binding, and malaria disease. *Curr Opin Microbiol* 9: 374-80

Kurtzhals JA, Goka BQ, Akanmori BD, Hviid L. 2001. The importance of strict patient definition in studies of malaria pathogenesis. *Trends Parasitol* 17: 313-4

Kyes S, Horrocks P, Newbold C. 2001. Antigenic variation at the infected red cell surface in malaria. *Annu Rev Microbiol* 55: 673-707

Kyes S, Taylor H, Craig A, Marsh K, Newbold C. 1997. Genomic representation of *var* gene sequences in *Plasmodium falciparum* field isolates from different geographic regions. *Mol Biochem Parasitol* 87: 235-8

Kyes SA, Kraemer SM, Smith JD. 2007. Antigenic Variation in *Plasmodium falciparum*: Gene Organization and Regulation of the *var* Multigene Family. *Eukaryot Cell*

Kyes SA, Rowe JA, Kriek N, Newbold CI. 1999. *Rifins*: a second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* 96: 9333-8

Kyriacou HM, Stone GN, Challis RJ, Raza A, Lyke KE, et al. 2006. Differential *var* gene transcription in *Plasmodium falciparum* isolates from patients with

cerebral malaria compared to hyperparasitaemia. *Mol Biochem Parasitol* 150: 211-8

Lavstsen T, Magistrado P, Hermsen CC, Salanti A, Jensen AT, et al. 2005. Expression of *Plasmodium falciparum* erythrocyte membrane protein 1 in experimentally infected humans. *Malar J* 4: 21

Lavstsen T, Salanti A, Jensen AT, Arnot DE, Theander TG. 2003. Sub-grouping of *Plasmodium falciparum* 3D7 *var* genes based on sequence analysis of coding and non-coding regions. *Malar J* 2: 27

Luginbuhl A, Nikolic M, Beck HP, Wahlgren M, Lutz HU. 2007. Complement Factor D, Albumin, and Immunoglobulin G Anti-Band 3 Protein Antibodies Mimic Serum in Promoting Rosetting of Malaria-Infected Red Blood Cells. *Infect. Immun.* 75: 1771-7

Mackintosh CL, Beeson JG, Marsh K. 2004. Clinical features and pathogenesis of severe malaria. *Trends Parasitol* 20: 597-603

Maitland K, Levin M, English M, Mithwani S, Peshu N, et al. 2003. Severe *P. falciparum* malaria in Kenyan children: evidence for hypovolaemia. *Qjm* 96: 427-34

Malaney P, Spielman A, Sachs J. 2004. The malaria gap. *Am J Trop Med Hyg* 71: 141-6

Marsh K, Forster D, Waruiru C, Mwangi I, Winstanley M, et al. 1995. Indicators of life-threatening malaria in African children. *N Engl J Med* 332: 1399-404

Mayor A, Bir N, Sawhney R, Singh S, Pattnaik P, et al. 2005. Receptor-binding residues lie in central regions of Duffy-binding-like domains involved in red cell invasion and cytoadherence by malaria parasites. *Blood* 105: 2557-63

Miller LH, Baruch DI, Marsh K, Doumbo OK. 2002. The pathogenic basis of malaria. *Nature* 415: 673-9

Miller LH, Good MF, Milon G. 1994. Malaria pathogenesis. *Science* 264: 1878-83

Ministry of Health Government of Tanzania NMCPT. 2003. *National Malaria Medium-Term Strategic Plan 2003-2007*

Molyneux ME, Taylor TE, Wirima JJ, Borgstein A. 1989. Clinical features and prognostic indicators in paediatric cerebral malaria: a study of 131 comatose Malawian children. *Q J Med* 71: 441-59

Montgomery J, Milner DA, Jr., Tse MT, Njobvu A, Kayira K, et al. 2006. Genetic analysis of circulating and sequestered populations of *Plasmodium falciparum* in fatal pediatric malaria. *J Infect Dis* 194: 115-22

Montgomery J, Mphande FA, Berriman M, Pain A, Rogerson SJ, et al. 2007. Differential *var* gene expression in the organs of patients dying of *falciparum* malaria. *Mol Microbiol* 65: 959-67

Mota MM, Pradel G, Vanderberg JP, Hafalla JC, Frevert U, et al. 2001. Migration of *Plasmodium* sporozoites through cells before infection. *Science* 291: 141-4

Newbold C, Warn P, Black G, Berendt A, Craig A, et al. 1997. Receptor-specific adhesion and clinical disease in *Plasmodium falciparum*. *Am J Trop Med Hyg* 57: 389-98

Newton CR, Krishna S. 1998. Severe *falciparum* malaria in children: current understanding of pathophysiology and supportive treatment. *Pharmacol Ther* 79: 1-53

Nielsen MA, Staalsoe T, Kurtzhals JA, Goka BQ, Dodoo D, et al. 2002. *Plasmodium falciparum* variant surface antigen expression varies between isolates causing severe and nonsevere malaria and is modified by acquired immunity. *J Immunol* 168: 3444-50

Nielsen MA, Vestergaard LS, Lusingu J, Kurtzhals JA, Giha HA, et al. 2004. Geographical and temporal conservation of antibody recognition of *Plasmodium falciparum* variant surface antigens. *Infect Immun* 72: 3531-5

Normark J, Nilsson D, Ribacke U, Winter G, Moll K, et al. 2007. PfEMP1-DBL1{alpha} amino acid motifs in severe disease states of *Plasmodium falciparum* malaria. *Proc Natl Acad Sci U S A* 104: 15835-40

Ofori MF, Staalsoe T, Bam V, Lundquist M, David KP, et al. 2003. Expression of variant surface antigens by *Plasmodium falciparum* parasites in the peripheral blood of clinically immune pregnant women indicates ongoing placental infection. *Infect Immun* 71: 1584-6

Peters J, Fowler E, Gatton M, Chen N, Saul A, Cheng Q. 2002. High diversity and rapid changeover of expressed *var* genes during the acute phase of *Plasmodium falciparum* infections in human volunteers. *Proc Natl Acad Sci U S A* 99: 10689-94

Peters JM, Fowler EV, Krause DR, Cheng Q, Gatton ML. 2007. Differential changes in *Plasmodium falciparum var* transcription during adaptation to culture. *J Infect Dis* 195: 748-55

Rasti N, Wahlgren M, Chen Q. 2004. Molecular aspects of malaria pathogenesis. *FEMS Immunol Med Microbiol* 41: 9-26

Reeder JC, Cowman AF, Davern KM, Beeson JG, Thompson JK, et al. 1999. The adhesion of *Plasmodium falciparum*-infected erythrocytes to chondroitin sulfate A is mediated by *P. falciparum* erythrocyte membrane protein 1. *Proc Natl Acad Sci U S A* 96: 5198-202

Roberts DJ, Pain A, Kai O, Kortok M, Marsh K. 2000. Autoagglutination of malaria-infected red blood cells and malaria severity. *Lancet* 355: 1427-8

Robinson BA, Welch TL, Smith JD. 2003. Widespread functional specialization of *Plasmodium falciparum* erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. *Mol Microbiol* 47: 1265-78

Rogerson SJ, Hviid L, Duffy PE, Leke RF, Taylor DW. 2007. Malaria in pregnancy: pathogenesis and immunity. *Lancet Infect Dis* 7: 105-17

Rottmann M, Lavstsen T, Mugasa JP, Kaestli M, Jensen AT, et al. 2006. Differential expression of *var* gene groups is associated with morbidity caused by

Plasmodium falciparum infection in Tanzanian children. *Infect Immun* 74: 3904-11

Rowe JA. 2005. Rosetting. In *Molecular approaches to malaria.*, ed. IW Sherman, pp. 416–26. Washington, DC: ASM Press

Rowe JA, Handel IG, Thera MA, Deans AM, Lyke KE, et al. 2007. Blood group O protects against severe *Plasmodium falciparum* malaria through the mechanism of reduced rosetting. *Proc Natl Acad Sci U S A*

Rowe JA, Kyes SA. 2004. The role of *Plasmodium falciparum var* genes in malaria in pregnancy. *Mol Microbiol* 53: 1011-9

Rowe JA, Moulds JM, Newbold CI, Miller LH. 1997. *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature* 388: 292-5

Rowe JA, Rogerson SJ, Raza A, Moulds JM, Kazatchkine MD, et al. 2000. Mapping of the region of complement receptor (CR) 1 required for *Plasmodium falciparum* rosetting and demonstration of the importance of CR1 in rosetting in field isolates. *J Immunol* 165: 6341-6

Russo CA, Takezaki N, Nei M. 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Mol Biol Evol* 13: 525-36

Sachs J, Malaney P. 2002. The economic and social burden of malaria. *Nature* 415: 680-5

Schellenberg D, Menendez C, Aponte J, Guinovart C, Mshinda H, et al. 2004. The changing epidemiology of malaria in Ifakara Town, southern Tanzania. *Trop Med Int Health* 9: 68-76

Scherf A, Carter R, Petersen C, Alano P, Nelson R, et al. 1992. Gene inactivation of Pf11-1 of *Plasmodium falciparum* by chromosome breakage and healing: identification of a gametocyte-specific protein with a potential role in gametogenesis. *Embo J* 11: 2293-301

Scherf A, Hernandez-Rivas R, Buffet P, Bottius E, Benatar C, et al. 1998. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in *Plasmodium falciparum*. *Embo J* 17: 5418-26

Smith JD, Chitnis CE, Craig AG, Roberts DJ, Hudson-Taylor DE, et al. 1995. Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* 82: 101-10

Smith JD, Craig AG, Kriek N, Hudson-Taylor D, Kyes S, et al. 2000a. Identification of a *Plasmodium falciparum* intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. *Proc Natl Acad Sci U S A* 97: 1766-71

Smith JD, Gamain B, Baruch DI, Kyes S. 2001. Decoding the language of *var* genes and *Plasmodium falciparum* sequestration. *Trends Parasitol* 17: 538-45

Smith JD, Subramanian G, Gamain B, Baruch DI, Miller LH. 2000b. Classification of adhesive domains in the *Plasmodium falciparum* erythrocyte membrane protein 1 family. *Mol Biochem Parasitol* 110: 293-310

Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, et al. 1995. The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* 82: 89-100

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596-9

Taylor HM, Kyes SA, Harris D, Kriek N, Newbold CI. 2000a. A study of *var* gene transcription in vitro using universal *var* gene primers. *Mol Biochem Parasitol* 105: 13-23

Taylor HM, Kyes SA, Newbold CI. 2000b. *Var* gene diversity in *Plasmodium falciparum* is generated by frequent recombination events. *Mol Biochem Parasitol* 110: 391-7

Treutiger CJ, Heddini A, Fernandez V, Muller WA, Wahlgren M. 1997. PECAM-1/CD31, an endothelial receptor for binding *Plasmodium falciparum*-infected erythrocytes. *Nat Med* 3: 1405-8

Treutiger CJ, Hedlund I, Helmby H, Carlson J, Jepson A, et al. 1992. Rosette formation in *Plasmodium falciparum* isolates and anti-rosette activity of sera from Gambians with cerebral or uncomplicated malaria. *Am J Trop Med Hyg* 46: 503-10

Treutiger CJ, Scholander C, Carlson J, McAdam KP, Raynes JG, et al. 1999. Rouleaux-forming serum proteins are involved in the rosetting of *Plasmodium falciparum*-infected erythrocytes. *Exp Parasitol* 93: 215-24

Trimnell AR, Kraemer SM, Mukherjee S, Phippard DJ, Janes JH, et al. 2006. Global genetic diversity and evolution of *var* genes associated with placental and severe childhood malaria. *Mol Biochem Parasitol* 148: 169-80

Tuikue Ndam NG, Fievet N, Bertin G, Cottrell G, Gaye A, Deloron P. 2004. Variable adhesion abilities and overlapping antigenic properties in placental *Plasmodium falciparum* isolates. *J Infect Dis* 190: 2001-9

Tuikue Ndam NG, Salanti A, Bertin G, Dahlback M, Fievet N, et al. 2005. High level of var2csa transcription by *Plasmodium falciparum* isolated from the placenta. *J Infect Dis* 192: 331-5

Udomsangpetch R, Sueblinvong T, Pattanapanyasat K, Dharmkrong-at A, Kittikalayawong A, Webster HK. 1993. Alteration in cytoadherence and rosetting of *Plasmodium falciparum*-infected thalassemic red blood cells. *Blood* 82: 3752-9

Uneke CJ. 2007. *Plasmodium falciparum* malaria and ABO blood group: is there any relationship? *Parasitol Res* 100: 759-65

Vogt AM, Barragan A, Chen Q, Kironde F, Spillmann D, Wahlgren M. 2003. Heparan sulfate on endothelial cells mediates the binding of *Plasmodium*

falciparum-infected erythrocytes via the DBL1alpha domain of PfEMP1. *Blood* 101: 2405-11

Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, et al. 2006. A *var* gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature* 439: 1004-8

Voss TS, Thompson JK, Waterkeyn J, Felger I, Weiss N, et al. 2000. Genomic distribution and functional characterisation of two distinct and conserved *Plasmodium falciparum var* gene 5' flanking sequences. *Mol Biochem Parasitol* 107: 103-15

Wahlgren M, Carlson J, Udomsangpetch R, Perlmann P. 1989. Why do *Plasmodium falciparumm*-infected erythrocytes form spontaneous erythrocyte rosettes? *Parasitol Today* 5: 183-5

Waller KL, Nunomura W, Cooke BM, Mohandas N, Coppel RL. 2002. Mapping the domains of the cytoadherence ligand *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) that bind to the knob-associated histidine-rich protein (KAHRP). *Mol Biochem Parasitol* 119: 125-9

Ward CP, Clottey GT, Dorris M, Ji DD, Arnot DE. 1999. Analysis of *Plasmodium falciparum* PfEMP-1/*var* genes suggests that recombination rearranges constrained sequences. *Mol Biochem Parasitol* 102: 167-77

Wernersson R, Pedersen AG. 2003. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res* 31: 3537-9

Whitten MM, Shiao SH, Levashina EA. 2006. Mosquito midguts and malaria: cell biology, compartmentalization and immunology. *Parasite Immunol* 28: 121-30

WHO. 1990. Severe and complicated malaria. . *Trans R Soc Trop Med Hyg* 84 Suppl 2: 1-65

WHO. 2000. Severe *falciparum* malaria. *Trans R Soc Trop Med Hyg* 94 Suppl 1: S1-90

Appendix 1: INPATIENT FORM

IDENTIFICATION

1	Date admitted (dd/mm/yy)	|__|__| / |__|__| / |__|__|
2	Hospital number	|__|__|__|__|__|__|__|__|
3	Study number	| __|__|__|__|/
4	Child's first name	|__|__|__|__|__|__|__|__|__|__|
5	Date of birth (dd/mm/yy)	|__|__| / |__|__| / |__|__|
	Age (Months)	| | |
7	Sex (1=male, 2=female)	|__|
8	Mother's first name	|__|__|__|__|__|__|__|__|__|__|
9	Father's first name	|__|__|__|__|__|__|__|__|__|__|
10	Family name of father	|__|__|__|__|__|__|__|__|__|__|
11	Village	|__|__|__|__|__|__|__|__|__|__|
12	Balozi	|__|__|__|__|__|__|__|__|__|__|
12(a)	Mwenyekiti wa Kitongoji	|__|__|__|__|__|__|__|__|__|__|

HISTORY

13	Fever (1=yes, 2=no / no days)	|__| / |__|__|
14	Cough (1=yes, 2=no / no days)	|__| / |__|__|
15	Dyspnoea (1=yes, 2=no / no days)	|__| / |__|__|
16	Child sucking / drinking 1=more than usual, 2=less than usual, 3=as usual)	|__|
17	Diarrhoea (1=yes, 2=no / no days) *If no, go to question 20*	|__| / |__|__|
17(a)	Have you given any oral treatment to this diarrhoea episode? (1=yes, 2=no)	|__|
17(b)	Which one (probe) (1=porridge, 2=plain water, 3=ORS, 4=others	|__|
18	Nº of stools in the last 24 hours	|__|__|
19	Stool type (1=Watery, 2= bloody, 3=mucoid)	|__|
20	Dysentery (1=yes, 2=no / no days)	|__| / |__|__|
21	Vomiting (1=yes, 2=no / no days)	|__| / |__|__|
22	Has the child fitted during this illness? (1=yes, 2=no) *If no, go to Q25*	|__|

23 Nº of seizures within the last 24 hours |__|__|

24 Other significant history _____

25 Has the child been seen at any health unit for that illness? (1=yes,2=no) |__|

26 If yes, where? _____

27 Treatment given:

A _____ |__|__ d _____ |__|__|

B _____ |__|__ e _____ |__|__|

C _____ |__|__ f _____ |__|__|

EXAMINATION

General

28 Height (cm) |__|__|__|

29 Weight (kg) |__|__| : |__|

30 Axillary temperature(ºC) |__|__| : |__|

31 Pallor (1=yes, 2=no) |__|

32 Jaundice (1=yes, 2=no) |__|

33 Skin rash (1=scabies, 2=other (mild), 3=other (severe), 4=none) |__|

34 Visible pus in ears ? (1= yes, 2 = no) |__|

35 Oral candidiasis (1=yes, 2= no) |__|

Respiratory & Cardiovascular

36 Pulse (timed over full minute) |__|__|__|

37 Respiratory rate (time over full minute) |__|__|__|

38 Nasal flaring (1=yes, 2=no) |__|

39 Indrawing (1=yes, 2=no) |__|

40 Crackles/creps/bronchial breathing (1= yes, 2=no) __|

41 Wheeze/ronchi (1=yes, 2=no) |__|

42 Gallop rhythm (1=yes, 2=no) |__|

43 Hepatomegaly (1=yes, 2=no /length cm) |__| / |__|__|

44 Splenomegaly (1=yes, 2=no /length cm) |__| / |__|__|

Nutritional & Hydration Status

45 Flaky paint skin (1=yes, 2=no) |__|

46 Orange hair (1=yes, 2=no) |__|

47 Visible wasting (1=yes, 2=no) |__|

48 Oedema (1=general, 2=face, 3=periphery, 4=other, 5 =none) |__|

49 Mucous membranes (1=moist, 2=dry, 3=very dry) |__|

50 Skin pinch goes back (1=quickly, 2=slowly, 3= very slowly) |__|

51 Level of consciousness (1=well/alert, 2=restless/irritable, 3=lethargic/unconscious)

52 Dehydration (none=1, some =2, severe=3) |__|

Neurological

53 Fontanelle (1=normal, 2=sunken, 3=bulging, 4=na) |__|

54 Position (1=decerebrate, 2=decorticate, 3=opisthotomia, 4=normal) |__|

55 Neck stiffness (1=yes, 2=no) |__|

56 Can the child sit down? (1=yes, 2=no) |__|

57 Eye movements (0=not directed, 1=directed) |__|

58 Verbal response (0=none, 1=inappropriate cry, 2=appropriate cry) |__|

59 Motor response (0=none, 1=withdraw from pain, 2=localises pain) |__|

60 Has the child had a convulsion in the last hour or anticonvulsant treatment in the last |__|

61(a) Provisional diagnosis by OPD Clinician |__|__|/|__|__|

61(b) Provisional diagnosis by IHRDC Clinician

LABORATORY AND COMPLEMENTARY INVESTIGATIONS

61 Microcapillary tube and blood slides sent to IC lab? (1=yes, 2=no) |__|

62 Sample brady number? (affix sticker here)

63 Blood glucose (mg/dl) |__|__|__|

64 Hospital PCV on admission (%) |__|__|

65 Initials of admitting officer |__|__|

AT DISCHARGE

66 Hospital parasitaemia on admission (1=positive, 0=negative) |__|

67 Chest x-ray taken (yes=1 ,no=2) |__|

68 If yes (normal=1 ,abnormal=2) |__|

69 LP done (1=yes, 2=no) |__|

70 If yes (normal=1,abnormal=2) |__|

71 Has the child received a blood transfusion during admission? (1=yes, 2=no) |__|

72 Date of transfusion |__|__| / |__|__| / |__|__|

73 What was the last hospital PCV recorded before transfusion? |__|__|

73(a) Hospital PCV on discharge |__|__|

74 Final diagnosis 1:_____ |__|__|__|

75 Final diagnosis 2:_____ |__|__|__|

75(a) Final diagnosis 3:_____ |__|__|__|

76 Date of discharge (dd/mm/yy) |__|__| / |__|__| / |__|__|

77 Outcome: (Alive=1, dead=2, absconded=3, 4=transferred) |__|

78 Treatment received during admission:

A _____ |__|__| d _____ |__|__|

B _____ |__|__| e _____ |__|__|

C _____ |__|__| f _____ |__|__|

Appendix 2: Analysed sequences

2.1 DBL-1α multiple alignments of predominant sequences from IAM isolates

```
F8IAM10   ARSFADIGDIVRGKDLFLG---------NTYESAQRDELEKNLKTIFEKIYNDVTNGSNA 51
F5IAM10   ARSFADIGDIVRGKDLYLGG--------NNKEKKRRDELEKNLKKIFEKIKGD--NNSKL 50
C4IAM11   ARSFADIGDIIRGKDIFIG-------------NKKKDKLEENLKTIFGKIYEKLNGAKEY 47
A8IAM5    ARSFADIGDIVRGRDLYSG-------NSKE---KKKRCLENKLKEIFGDIYKDVTKNNR- 49
A10IAM11  ARSFADIGDIVRGKDLYLG-------NKKKNQARKEKELENKLKEYFEKIYNSLES---- 49
A3IAM12   ARSFADIGDIIRGKDLYLGKKKK-QNETKTE--TEREKLEQKLKEIFAKIHSEVTSTSGN 57
F2IAM18   ARSFADIGDIIRGKDLYLGKKKKKQNETKTE--TEREKLEQKLKEIFAKIHSEVTSTSGN 58
D5IAM5    ARSFADIGDIIRGKDLYLEKKKKKQNETKTE--TEREKLEQKLKEIFAKIHSEVTSTSGN 58
E1IAM7    ARSFADIGDIVRGKDLYLG------YDEKEK--EQRKKLEKNLKEIFTQIYNDVT-TNGK 51
F7IAM17   ARSFADIGDIVRGKDLYLG------YDDKEK--KRREKLENKLKEISTQIHEDVT-KNG- 50
E2IAM18   ARSFADIGDIIRGKDLYLG-------NPQES--AQRIILEKNLKKYFQQIHNDVT-SSGR 50
F10IAM10  ARSFADIGDIIRGKDLFIGYD--------EKDRKEKKQLQQNLKNIFGKIHSEVTNG-SN 51
H6IAM18   ARSFADIGDIVRGKDLYIGN----------RKEKEKEELQKNLKSIFKKIYGELKNGKTN 50
F3IAM17   ARSFADIGDIVRGKDLFYGN---------PQEKKQRKELDKKLKEVFGKIHEGLKNG--- 48
H8IAM7    ARSFADIGDIVRGRDLYFGKRKK---KNQKETETERKKLEKNLKDIFKKIYGELSRTSTN 57
A5IAM12   ARSFADIGDIIRGKDLYLG-------DNRKDRE-QKVKLENKLKEIFAKIHENLGTQDAI 52
A7IAM11   ARSFADIGDIIRGKDLYRG--------NSKEKD----NLEKKLIEYFQKIHGGL-TGDAQ 47
A2IAM12   ARSFADIGDIIRGKDLYLDH------------EPGKQHLEERLETMFQNIQYNNTELKDI 48
F12IAM17  ARSFADIGDIIRGKDLYLDH------------EPGKQHLEERLETMFQNIQYNNTELKNI 48
F6IAM7    ARSFADIGDIVRGKDLFLGH------------KQGKQKLEASLKTMFQNILSTIDQLKRL 48
D4IAM5    ARSFADIGDIVRGTDMFLGS------------NKEKGKIENSLQNIFKNIKKNNKKLKDL 48
          **********:** *::                      ::  *     .*

F8IAM10   E--AAKERYKGDADNNYFQLREDWWTANRATIWEAITC--DVHGSDYFRPTCS-MNGSGA 106
F5IAM10   S--TLKD----D------QIREYWWALNRETVWEAMTC--KAEG-AYFHATCS-MNGSGA 94
C4IAM11   Y--QDKN------GGNFLKLREDWWALNRNDVWKAMTCSEDLKNSSYFHATCIDRKGSCS 99
A8IAM5    ---ALQARYT-DNKN-YYKLREDWWDANRETVWKAITCGSA--GGKYFRKTCAGG---TS 99
A10IAM11  ---SIKSQYD-DEANNFFKLREDWWALNREEVWKAITCGHP--DGTYFRNTCAGG---TT 100
A3IAM12   NKEVLKARYD-GDGDNYYQLREDWWDANRLDVWKAITCGAP-HGAQYFRQTCNDDG-TSS 114
F2IAM18   NKEVLKARYD-GDGDNYYQLREDWWDANRLDVWKAITCGAP-HGAQYFRQTCNDDG-TSS 115
D5IAM5    NKEVLKARYD-GDGDNYYQLREDWWDANRLDVWKAITCGAP-HGAQYFRQTCNDDG-TSS 115
E1IAM7    NVDALKTRYN-DDTPDFFKLREDWWDANRETVWKALTCDHRLGGSQYFRPTCGSKAKTAT 110
F7IAM17   ----AEARYN-DDTGDFFQLREDWWTANRSTIWEAITCNAG-GGNRYFRQTCGS----GN 100
E2IAM18   N--GVKDRYK-DTDKNFFQLREDWWYANRETVWKAITCNAQ--GFDYFRQTCGDDEKTAT 105
F10IAM10  A-EAAKARYK--DTTDFYQLREDWWDANRETVWEAITCG-AAGG-TYFRATCSDEENKST 106
H6IAM18   G-EAAKVHYQE-DGQNYYKLREDWWTANRETVWEAITCN-AGGG-TYFRGTCGKND---T 103
F3IAM17   ---KAKERYK--DTTNYYQLREDWWTANRETVWKAITCA-AKVGDTYFMESRTNSY---K 99
H8IAM7    GRNGELQERYEDKDGNYLKLREDWWTANRDQVWKAITCDAR-DNAQYFRATCAGEQ---- 112
A5IAM12   G-------HYEDAKKNYYKLREDWWTANRGTVWKAITCGAG-KHDKYFRKTCNGGS---- 100
A7IAM11   T-------HYNDKSGNFFKLREDWWDANRQEIRNAIICDVP-EDAKYLEQSDGSQSG--- 96
A2IAM12   P---------------LPKVREYWWALNRDQVWKAITCHAG-KDDAYFRNSSGGEY---K 89
F12IAM17  P---------------LPKVREYWWALNRDQVWKAITCHAG-KDDAYFRNSSGGEY---K 89
F6IAM7    S---------------INAVREYWWEIKRQEVWKAITCSAG-EDDTYSKYLGDRTT---G 89
D4IAM5    T---------------DKQIREYWWALNRKEVWKALTCSVP-YEAYYFTYKSDNFR---T 89
                          :** **  :*  : :*: *       *
```

```
F8IAM10   QANNKCRCTKSSGAK-DGEVNIVPTYFDYVPQYLRWFEEWAEDL 149
F5IAM10   QANNKCRCDKEKAGKGSGDVTIVPTHFDYVPQFLRWFEEWAEDL 138
C4IAM11   QAKDKCRCKDEKGTN----TDQVPTYFDYVPQYLRWFEEWAEDL 139
A8IAM5    LTEDNCRCAANI---------DPPTYFDYVPQYLRWFEEWAEDL 134
A10IAM11  PTPNKCRCATN----------DAPTNFDYVPQYLRWFEEWAEDL 134
A3IAM12   RAIHQCRCQKKDGTH---DSDQVPTYFDYVPQYLRWFEEWAEDL 155
F2IAM18   RAIHQCRCQKKDGTH---DSDQVPTYFDYVPQYLRWFEEWAEDL 156
D5IAM5    RAIHQCRCQKKDGTH---DSDQVPTYFDYVPQYLRWFEEWAEDL 156
E1IAM7    QTPSQCRCSDN----------QVPTYFDYVPQFLRWFEEWAEDL 144
F7IAM17   WTKDKCRCDDKP-------NTDPPTYFDYVPQYLRWFEEWAEDL 137
E2IAM18   RVKDKCRCDGD----------QVPTYFDYVPQFLRWFEEWAEDL 139
F10IAM10  LASNKCRCAGKNA-------DQVPTNFDYVPQYLRWFEEWAEDL 143
H6IAM18   WTREDCRCDGSN---------VVPTYFDYVPQYLRWFEEWAEDL 138
F3IAM17   FSGDKCGHNDDN----------VPTNLDYVPQYLRWFEEWAEDL 133
H8IAM7    RTKGYCRCNGDKPDN-DMVNIDPPTYFDYVPQYLRWFEEWAEDL 155
A5IAM12   PTKGYCRCNGDQPND-HKANIDPPTHFDYVPQFLRWFEEWAEDL 143
A7IAM11   SHQTKCRCHSG----------SVLTNFDYVPQYLRWFEEWAEDL 130
A2IAM12   FTSGYCGRNEGK----------VPTNLDYVPQHLRWFEEWAEDL 123
F12IAM17  FTSGYCGRNEGK----------VPTNLDYVPQHLRWFEEWAEDL 123
F6IAM7    VSHGQCGHMDEN----------VPTHFDYVPQFLRWVEEWAEDL 123
D4IAM5    FSGYWCGHYEGA----------PPTNLDYVPQFLRWFEEWAEDL 123
```

2.2 DBL-1α multiple alignments of predominant sequences from SM isolates

```
A1ISM49   ARSFADIGDIVRGRDLFLGNTY-------ESAQRDQLDKK----LKEIFTQIYNDVTTN- 48
C4ISM16   ARSFADIGDIIRGRDLFYGNTQ-------EKTKRKQLDKK----LKDIFGDIYKELRKN- 48
D2ISM49   ARSFADIGDIVRGKDPFYGNPD-------EIKQRQQLEDK----LKEIFQKIHDKLDS-- 47
A3ISM11   ARSFADIGDIVRGRDIFRGNDE-------EKKKRDELDDK----LKKIFGKIHGGLTG-- 47
A1ISM16   ARSFADIGDIIRGKDLYLGYDDE------EKSRRKQLDDK----LKDIFKEIHDDVTTN- 49
E5ISM48   ARSFADIGDIVRGKDLYLGYDDK------EKKRREKLENK----LKEIFTQIHEDVTKN- 49
E2ISM3    ARSFADIGDIIRGKDLYLGYDQK------EKDRRDELEKN----LKTIFGKIHSDVTSG- 49
A5ISM49   ARSFADIGDIIRGKDLYLGKKKKQNETKTETEREKLEQK----LKEIFAKIHSEVTSTS 56
ISM11F10  ARSFADIGDIIRGKDLFIGNNK-----------RDKLEKQ----LKEYFKNIYDNLN--- 42
F1ISM3    ARSFADIGDIIRGKDLFRGYNEK------DRKEKEQLQDS----LKNIFKEIYDDVTSDK 50
B5ISM11   ARSFADIGDIIRGKDLFRGYDYE------EKNRREQLEEN----LKTIFGNIYEELREEQ 50
G12ISM3   ARSFADIGDIIRGKDLYLDHEPG----------KQHLEER----LERIFENIKK------ 40
A10ISM2   ARSFADIGDIIRGKDLHLDHEPG----------KQHLEER----LERIFENIKK------ 40
ISM51C11  ARSFADIGDIIRGKDLYLDHEPG----------KQHLEER----LERIFANIQK------ 40
F9ISM33   ARSFADIGDIIRGKDLYLDHEQG----------NNRLEAR----LKTIFQNIQN------ 40
E8ISM51   ARSFADIGDIIRGKDLFLGHEQR----------KKYLEAR----LEAMFDNIKK------ 40
H9ISM33   ARSFADIGDIIRGKDLFLGHEQR----------KKYLEAR----LEAMFDNIKK------ 40
G5ISM48   ARSFADIGDIVRGTDMFLGSNKE----------KEKIENS----LQNIFKNIKK------ 40
C9ISM2    ARSFADIGDIVRGKDLYLGDKGE----------KKKLEKN----LKDIFKQIHEKLTDP- 45
A1ISM2    ARSFADIGDIIRGKDLFLGHKEQ----------KKKLQEN----LEKIFNKFKT------ 40
E6ISM51   ARSFADIGDIVRGRDMFL-PNKD----------DKVQKG-----LQVVFEKINNGLKKI- 43
F8ISM33   ARSFADIGDTVKGKDMFK-PNDA----------DKVEKG-----LQVVFGKIYKSLPSP- 43
C1ISM16   ARSFADIGDIVRGKDMFK-RNEE----------DAVQKG-----LRAVFKKINDNLKVK- 43
E9ISM48   ARSFADIGDIVRGKDLFLGNNDN----------DKVKKEKLQNNLKSIFAKIYKELKLE- 49
          ********* ::* * .                    .      *.  * .:
```

```
A1ISM49  -------G-KKPALQKRYKKDGKDPDFFKLREDWWYANRQEIWKAITCKV----ENAQYF 96
C4ISM16  -------G-KKGELQKRYQKDG-DKDFFQLREDWWEENRETVWKAITCDAP---PDAQYF 96
D2ISM49  ------------KAQARHQHDAP--DYYQLREDWWTANRATVWKAITCDDDKKLASASYF 93
A3ISM11  ------------DAQTHYQDDAKK-NFYQLREDWWTANRATIWEAITCKAD---TGNAYF 91
A1ISM16  -------GKNWQTLKKRYEGDTD-GNYYKLREDWWTANRATIWEALTCDDDNKLAGAHYF 101
E5ISM48  -------G-----AEARYNDDT--GDFFQLREDWWTANRSTIWEAITCNAG---GGNRYF 92
E2ISM3   -------K-NAEELKARYNGDKN-NDFFKLREDWWEANRETIWRAITCGHP----GGTYF 96
A5ISM49  -------GNNKEVLKARYDGDGD--NYYQLREDWWDANRLDVWKAITCGAP---HGAQYF 104
ISM11F10 -------GAQKHYSDD--DKGTK--NYYQLREDWWALNRQEIWKALTCES----GGGRYF 87
F1ISM3   N------GKNAEELKERYGQDSP--NFYKLREDWWNANRLDVWKAITCKAE----NAQYF 98
B5ISM11  TKRKRAKPKNGQALQARYKKDGD--NFFKLREDWWYANRLEVWKAITCHAG---KDDAYF 105
G12ISM3  ---------------KNNNNELNNLSLDKFREYWWALNRVQVWKAITCRAE---EKDIYS 82
A10ISM2  ---------------KNNNNELNNLSLDKFREYWWALNRVQVWKAITCRAE---EKDIYS 82
ISM51C11 ---------------ENG--DINTLKPEEVREYWWALNRVQVWKAITCRAE---EKDIYS 80
F9ISM33  ---------------KNKS-PLDKLSLDKFREYWWALNRVQVWKAITCGAT---MNDIFS 81
E8ISM51  ---------------NNKK-QLGELSTAQVRGYWWALNRDQVWKAITCGAT---MNDISF 81
H9ISM33  ---------------NNKK-QLGELSTAQVREYWWALNRGQVWKAITCGAT---MNDISF 81
G5ISM48  --------------NNKKLKDLTDK---QIREYWWALNRKEVWKALTCSVP---YEAYYF 80
C9ISM2   -------------RAKDHYKDEKDGNLFQLREDWWTANRDQVWKAITCNAP---YKAWYF 89
A1ISM2   -----------------IYAGLEDVAIDDIREYWWALNREDVWKAITCKAP---TGADYF 80
E6ISM51  -------------GINAYNDG--SGNYSKLREVWWNVNRDQVWRAITCSAP---GDVNYF 85
F8ISM33  -------------AKKHYPDGDKSGNYVKLREDWWTANRDQVWRAITCSAP---GDVNYF 87
C1ISM16  -------------EISDYDND---PNYYKLREDWWNVNREQVWRAITCYIP---YYVNYF 84
E9ISM48  -------------KNSDYKDDDIDGNYYKLREDWWTVNRDQVWKALTCFAD---GSEEYF 93
                                  ..*  **  **  :*.*:**

A1ISM49  KDTCSTG-----GHYEKCRCNGDQPKSGKGGDVNIVPTYFDYVPQYLRWFEEWAEDL 148
C4ISM16  RGTCGDNEKTATQTPSQCRCNDDQ----------VPTYFDYVPQYLRWFEEWAEDL 142
D2ISM49  RETCGGDGKTGTQAKRQCRCEDAN--------ANQVPTYFDYVPQYLRWFEEWAEDL 142
A3ISM11  RATCGGSGETPLVTPSQCRCKKNDG----GDDADQVPTYFDYVPQYLRWFEEWAEDL 144
A1ISM16  RQTCGSNAKSATQASNKCRCSDKVN-TDPP-------TYFDYVPQFLRWFEEWAEDL 150
E5ISM48  RQTCGS----GNWTKDKCRCDDKPN-TDPP-------TYFDYVPQYLRWFEEWAEDL 137
E2ISM3   RKTCVG----ENETQNNCRCDDKPK-SGKPREVNIVPTYFDYVPQYLRWFEEWAEDL 148
A5ISM49  RQTCND-DGTSSRAIHQCRCQKKDG----THDSDQVPTYFDYVPQYLRWFEEWAEDL 156
ISM11F10 RETC---AGGTSRTQDDCRCRTND----------VPTYFDYVPQYLRWFEEWAEDL 130
F1ISM3   RNACSKG---TTPTNEKCHCIDET----------VPTYFDYVPQYLRWFEEWAEDL 141
B5ISM11  RKSGYR-----EFEFTDGHCGNKDG---------TVPTNLDYVPQHLRWFEEWAEDL 148
G12ISM3  KTTDN----GKLLLWNY-NCGHHVNK--------DVPTNLDYVPQFLRWFEEWAEDL 126
A10ISM2  KTTDN----GKLLLWNY-NCGHHVNK--------DVPTNLDYVPQFLRWFEEWAEDL 126
ISM51C11 RIA------GDTTIWND-NCGHHVNQ--------DVPTNLDYVPQYLRWFEEWAEDL 122
F9ISM33  KNIRN----SRTTLFYY-KCGHYVYK--------DVPTNLDYVPQFLRWFEEWAEDL 125
E8ISM51  KNIGN----GKLLLWNE-KCGHGDYN---------LLTNLDYVPQFLRWFEEWAEDL 124
H9ISM33  KNIGN----GKLLLWNE-KCGRGDYN---------LLTNLDYVPQFLRWFEEWAEDL 124
G5ISM48  TYKSD----NFRTFSGY-WCGHYEG---------APPTNLDYVPQFLRWFEEWAEDL 123
C9ISM2   MHSED----NKQLFSDY-KCGHYEG---------SPLTNLDYVPQFLRWFEEWAEDL 132
A1ISM2   VYKSG----SLLNFSSD-RCGHNNND--------GPLTNLDYVPQFLRWFEEWAEDL 124
E6ISM51  RKISG----DTRTFENAGKCRRHDN---------KVPTNLDYVPQFLRWFEEWAEDL 129
F8ISM33  RKESD----GSYVFSNRGPCGRNET---------DVPTNLDYVPQYLRWFEEWAEDL 131
C1ISM16  KKTSD----NTIVFTNDGKCGHYEG---------APPTNLDYVPQFLRWFEEWAEDL 128
E9ISM48  IQSEN----NTQLFSNP-KCGHEQG---------TVPTYLDYVPQFLRWFEEWAEDL 136
                          *                * :*****.**********
```

2.3 upsA multiple sequence alignment of predominant sequences from 6 isolates

```
ISM51upsAG3 TTTATTATATTTGTTGTAGGTGAAAAATATATGAATTTTACAGGAATATAAGGGTTTATT 60
ISM49upsAF3 TTTATTACATTTGTTGTAGGTGAAAAATATATGAATTTTACAGGAATATAAGGGTTTATT 60
ISM16upsAG6 TTTATTATATTTGTTGTAGGTGAAAAATATTCGATTTTTAATCGAATAATAGGGTTTATT 60
ISM11upsAH5 TTTATTATATTTGTTGTAGGTGAAAAATATTCGATTTTTAATTGAATATTAGTGTTTATT 60
ISM2upsAB12 TTTATTATATTTGTTGTAGGTGAAAAATATGTAAATTTTAATTGAATATTAATCTTTATT 60
ISM16upsAE10TTTATTATATTTGTTGTAGGTGAAAAATATTCGATTTTTAATCGAATAATAGGGTTTATT 60
ISM3upsAC6  TTTATTACATTTGTTGTAGGTGACAAATATATGAATTTTACAGGAATATAAGGGTTTATT 60
ISM49upsAF8 TTTATTACATTTGTTGTAGGTGACAAATATATGAATTTTACAGGAATATAGGGGTTTATT 60
IAM17upsAA3 TTTATTACATTTGTTGTAGGTGAAAAATATATGAATTTTACTGGAATATAAGTGTTTATT 60
IAM17upsAD3 TTTATTATATTTGTTGTAGGTGAAAAATATATGAATTTTACAGGAATATAAAGGTTTATT 60
ISM49upsAG3 TTTATTATATTTGTTGTAGGTGAAAAATATATGAATTTTACAGGAATATAAAGGTTTATT 60
ISM11upsAG9 TTTATTATATTTGTTGTAGGTGACAAATATATGAATTTTACAGGAATATAAGGGTTTATT 60
ISM3upsAB11 TTTATTATATTTGTTGTAGGTGACAAATATATGAATTTTACAGGAATATAAGGGTTTATT 60
ISM3upsAD6  TTTATTACATTTGTTGTAGGTGATGAATGTATGGATTTTACATGAATACAGTTGTTTATT 60
ISM2upsAD10 TTTATTATATTTGTTGTAGGTGAGAAATATATGAATTTTACATGAATATGAGGATTTATT 60
ISM16upsAE3 TTTATTATATTTGTTGTAGGTGAAAAATATGTGAATATTACATGAATAACAGGGTTTATT 60
ISM2upsAD3  TTTATTATATTTGTTGTAGGTGAAAAATATTCGATATTTATTTGAATATTAGGGTTTATT 60
IAM17upsAB7 TTTATTATATTTGGTGTAGGTGAAAAATATTCGATTTTTAATCGAATAATAGGGTTTATT 60
            ******* ***** *********  *** *       ***   *****      ******

ISM51upsAG3 AACAATATATAATAATAATGATTATTAGCATGAAAAAATAATAT-TGTTTTCACATAATA 119
ISM49upsAF3 AACAATATATAATAAGAATGATTATTAGCATGAAAAAATAATAT-TGTTTTCACATAATA 119
ISM16upsAG6 TACAATATTTGATTGTCATATTTATTACTATGAAGAACCAATAC-TGTTTTCACCATATA 119
ISM11upsAH5 TACAATATTTGATTGTCATGTTTATTACTATGAAGAACTAATAC-TGTTTTCACCATATA 119
ISM2upsAB12 AACAATATATATTAATAATGTTTATTACCATGAAGAATTAATAT-GGTTTTCACATAATA 119
ISM16upsAE10TACAATATTTGATTGTCATATTTATTACTATGAAGAACTAATAC-TGTTTTCACCATATA 119
ISM3upsAC6  AATTATATATAATAATCATGTTTATTACCATGAAAAAACAATAT-TGCTTCCACAATGTA 119
ISM49upsAF8 AATTATATATAATAATCATGTTTATTACCATGAAAAAACAATAT-TGCTTCCACAATGTA 119
IAM17upsAA3 AACAATATATTATAATAATGTTTATTACTATGAAAAAATAATAT-TGTTTTCACATAATA 119
IAM17upsAD3 AACAATATATAATAATCATGATTATTACCTTGGAAAAATTATAT-TGTTTTCGCAATATA 119
ISM49upsAG3 AACAATATATAATAATCATGATTATTACCTTGGAAAAATTATAT-TGTTTTCGCAATATA 119
ISM11upsAG9 AATTATATATAATAATCATGTTTATTACCATGAAAAAACAATAT-TGCTTCCACAATGTA 119
ISM3upsAB11 AATTATATATAATAATCATGTTTATTACCATGAAAAAACAATAT-TGCTTCCACAATGTA 119
ISM3upsAD6  AACAATATGTAATAATCATATTTATTACCATGAAGAAATAATAT-TGTTTTCACGATATA 119
ISM2upsAD10 AACATTATTTAATAATCATGGTAATTACCATGAAGAAATAATATATTTTTTAACAATATA 120
ISM16upsAE3 AATATTATATCATTACCATATTTATTAGTAGGAAAAAATAATGT-AGTTTTGAGAATATA 119
ISM2upsAD3  AACAATATCTGATTATAATGTTTATTGCGTTTAAGAAATAATAC-TGTTTTTACAATATA 119
IAM17upsAB7 TACAATATTTGATTGTCATATTTATTACTATGAAGAACTAATAC-TGTTTTCACCATATA 119
             *   *** *  *    **  * ***       * **   **      **        **

ISM51upsAG3 ATGATAGTTTTTGT----AAAATATATTTAGAGGATATATATTGGTTATAATAATTGTAA 175
ISM49upsAF3 ATGATAGTTTTTGT----AAAATATATTTAGAGGATATATATTGGTTATAATAATTGTGA 175
ISM16upsAG6 ATGTTAATATTTTT----AAAATATATTTAAAGGAAATGAATTGGTTATAATAACTGTAA 175
ISM11upsAH5 ATGTTAATATTTTT----AAAATATATTTAGAGGAAATGAATTGGTTATAATAATTGTGA 175
ISM2upsAB12 ATGATAATTTTTGT----AAAATATATTTAGAGGATATATATTTGTTATAATAATTGTAA 175
ISM16upsAE10ATGTTAATATTTTT----AAAATATATTTAAAGGAAATGAATTGGTTATAATAACTGTAA 175
ISM3upsAC6  ATAATAATATATTT----AAAATATATTTAGTGCATATATATTGATTATAATAATTGTTA 175
ISM49upsAF8 ATAATAATATATTT----AAAATATATTGAGATTGTATATATTGATTATAATAATTGTTA 175
IAM17upsAA3 ATGATAGTTTTTGT----AAAATATATTTAGAGGATATATATTGGTTATAATAATTGAAA 175
IAM17upsAD3 ATAATAATATATTT----AAAATATATTTAGGGGACATAAATTGATTATAATAATTGCAA 175
ISM49upsAG3 ATAATAATATATTT----AAAATATATTTAGGGGACATAAATTGATTATAATAATTGCAA 175
ISM11upsAG9 ATAATAATATATTTTTTTGAAATATATTTAGTGCATATATATTGATTATAATAACTGAAA 179
ISM3upsAB11 ATAATAATATATAT----AAAATATATTGAGATTGTATATATTGATTATAATAATTGTTA 175
ISM3upsAD6  ATCATAATGGTTTT----AAACTATTTTTAGAGGATATTTATTGATTACAATAATTTTAA 175
ISM2upsAD10 ATAATCATATTTTA----AAAATATGTTTAGAGGAAATATATTGGTTAAATTAATTATAA 176
ISM16upsAE3 ATAGTACTATAATT----AAAATATATTTAGAAGAAATATATGGGTTATAATAATTATAA 175
ISM2upsAD3  ATGATAATGATTTT----AAAATATATTTAGCGGAAACGAATTGGTTATAATAATTGTAA 175
IAM17upsAB7 ATGTTAATATTTTT----AAAATATATTTAAAGGAAATGAATTGGTTATAATAACTGTAA 175
            **  *  *           ** *** ** *      *   **   *** * *** *   *
```

```
ISM51upsAG3 TGTATTAGTATAAATTATTTAAATACTTTGTTAT--AAAAATGGGGACAGGTTCATCA-- 231
ISM49upsAF3 CGTTATATTATAGATTATTTTAATAGTTTGTTAT--TAAAATGGGGGGAGGACATACA-- 231
ISM16upsAG6 TCTTATAGTATAGATGTTTTATACATTGCGTTAT--AAAAATGGGGTCACAACCATCA-- 231
ISM11upsAH5 CGTTATATTATAGATTATTTTAATAGTTCGTTAT--AAAAATGGGGACAGGTTCATCA-- 231
ISM2upsAB12 TGTTATAGTACAGATGTTTTATATATTACGTTAT--AAAAATGGGGACAGGTTCATCA-- 231
ISM16upsAE10TCTTATAGTATAGATGTTTTATACATTGCGTTAT--AAAAATGGGGTCACAACCATCA-- 231
ISM3upsAC6  TGTATTAATATATATTATTTAAATATTTTATTAT--AAAAATGGGGAATACAAAATCA-- 231
ISM49upsAF8 TGTATTAATATATATTATTTAAATATTTTATTAT--AAAAATGGGGAATACACAATCATT 233
IAM17upsAA3 TGTCTTAATATAGATTATTAAAATATTTTATAAT--CAAAATGGGAGCATCACAATCA-- 231
IAM17upsAD3 TGTATTAATACATAATATTTAAATATTTTATTAT--AAAAATGGGGAATACAGAATCATC 233
ISM49upsAG3 TGAATTAATACATAATATTTAAATATTTTATTAT--AAAAATGGGGAATGAACAATCATC 233
ISM11upsAG9 TGTCTTAGTATAAATTATTTAAACATTTTGTTAT--AAAAATGGGGAATGCAACATCATC 237
ISM3upsAB11 TGTATTAATATATATTATTTAAATATTTTATTAT--AAAAATGGGGGATACAAAATCA-- 231
ISM3upsAD6  TGTCTTAGTATAGATTATTAAAATATTTTATAAT--CAAAATGGGATCATCGCATTCA-- 231
ISM2upsAD10 TATCATAGTATAGGATATTTAAATAGATTGATATTAAAAAATGGGATCACAAGAATCT-- 234
ISM16upsAE3 TGTTATAGTATACATTATTTAAATTTTTTAATAT--CAAAATGGGGGGAAATTCTTCA-- 231
ISM2upsAD3  TGTTATAGTATAGATGTTTTAAATATTACATTAT--AAAAATGGGGTCACAATCATCA-- 231
IAM17upsAB7 TCTTATAGTATAGATGTTTTATACATTGCGTTAT--AAAAATGGGGTCACAACCATCA-- 231
                 ** ** *     **   *          **   ********           *

ISM51upsAG3 --ACTCCT----TCGGTTCCAAAAGAT---GTTAAAAATGAAAGTCACAACAGTGCCAGA 282
ISM49upsAF3 --AAGCAT-------GTTCAAGTTGTT---GACATAAAGGAAAGTCCTAACAGCGCTAGA 279
ISM16upsAG6 --AAACCT----TCGAAACCAAGTGTT---GATACAAATGAAAGTCACAAAAGTGCCCGA 282
ISM11upsAH5 --ATTCCT----TCGGTTCCAAAAGAT---GTTAAAAATGAAAGTGAGAAAAGTGCCAGA 282
ISM2upsAB12 --ACTCCT----TCGGTTCCAAAAGAT---GTTACAAATGAAAGTCACAAAAGTGCCAGA 282
ISM16upsAE10--AAACCT----TCGAAACCAAGTGTT---GATACAAATGAAAGTCACAAAAGTGCCAGA 282
ISM3upsAC6  --ACGCCT----TCGGTTCCAAAAGAT---GTTAAATATGAAAGTCACAAAAGTGCCAGA 282
ISM49upsAF8 AAATGAAG----AGGAGGCTAAAACCCCTAGTTTAACA-GAAAGTCACAACAGTGCCAGA 288
IAM17upsAA3 --AAACCT----TCGAAACCAAGTGTT---GATACAAATGAAAGTCACAACAGTGCAAGG 282
IAM17upsAD3 ATTAGAG-----GGAGAGGCTAGAAGCCCTAGTATAATAGAAAGTGAGAACAGTCCAAGA 288
ISM49upsAG3 ATCATCA-----GAGGGGGATAAAATTCCTAGTATAATAGAAAGTGAGAACAGTCCAAGA 288
ISM11upsAG9 AACATCATCAGAGGGGGAGATGAAATCCCTTTTATAAAAGATAGTGAGAACAGTCCGAGA 297
ISM3upsAB11 --ACGCCT----TCGGTTCCAAAAGAT---GTTAAATATGAAAGTCACAACAGTGCCAGA 282
ISM3upsAD6  --ACAAAT----GATACTAAAAGTCCT---ACTCTAAGTGAAAGTCACAAGAGTGCCAGA 282
ISM2upsAD10 --AAACCT----GTGGATACAAGCGAT---GTTAAAAATGAAAGTCACAACAGTGCTAGA 285
ISM16upsAE3 --AAAGGT----GCTCCTACATACTAT---AC---AAATGAAAGTGAAAAAAGTGCCAGA 279
ISM2upsAD3  --AAACCT----TCGGAACCTAGTGTT---GATACAAATGAAAGTCACAAAAGTGCCAGA 282
IAM17upsAB7 --AAATCT----TTGGAACCAATTGTT---GATACAAATGAAAGTCACAAAAGTGCCAGA 282
                                          *   ** ***   ** **  *  *

ISM51upsAG3 AATGTTTTGGAAAATATTGGAAAATGGATAAAAGATGAGATTAATAAAAATAGTAAT--- 339
ISM49upsAF3 AATGTTTTGGAAAATATTGGAATAGAAATATATAATGAGGAAAAAAAAAAAGTGAATGGA 339
ISM16upsAG6 AATGTTTTGGAAAATATTGGAATAGAAATATATAAGGAAATAGAAAAAACGATTCCA--- 339
ISM11upsAH5 AATGTTTTGGAAAATATTGGAAAATGGATAAAAGATGAGAGAGTAAAATCGAGTAAA--- 339
ISM2upsAB12 AATGTTTTGGAAGAAATTGGAAAAAAGATAAAAGATAAGACAGAAAAAGGGAATAAC--- 339
ISM16upsAE10AATGTTCTGGAAAATATTGGAAGCGCCATAAAAGATAAGAGGCAAATTGAGAGTAAA--- 339
ISM3upsAC6  AATGTTCTGGAAAAAATTGGACTAGAAATATATAA---GGAAATAGAAAAAACGATTCCA 339
ISM49upsAF8 AATGTTTTGGAAAATATTGGAATAGGAATATATAATGAGGAAAAAAAAAATGTAAATAAA 348
IAM17upsAA3 GGTGTTTTGGAAGAAATTGGAAAAAAGATAAAAGATAAGACAGAAAAAGAGAGTAAA--- 339
IAM17upsAD3 AATGTTTTGGAACGTTATGCCAAAAATATAAG---ACAGGCATCAAAAGCTGAAAATGAA 345
ISM49upsAG3 AATGTTTTGGAACGTTATGCCAAAAATATAAG---ACAGGCATCAAAAGATGAAAATGAA 345
ISM11upsAG9 AATGTTTTTGAACGTTATGCCGAACATATAAAGAAACAGGCAGAAAATGATGCAAAAAAA 357
ISM3upsAB11 AATGTTTTGGAAGATATTGGTAAAGGAATAAAAGATAAAGTAACAAAGGATGCAGAAAAA 342
ISM3upsAD6  AATGTTCTGGAAAACATTGCCCAAGATATAAGGAAACAAGTGCAAAAAGATGCACAAAAA 342
ISM2upsAD10 AATGTTTTTGAATACATTGCAGAAATAATAAGCAACGAGGTAAAAGAAAATGCTGAAAAA 345
ISM16upsAE3 AACGTTTTGGAAAATTTCGCAAAAGATATAAAAGGAAAAGCGTCAAATGACGCAAATAGG 339
ISM2upsAD3  AATGTTTTGGAAAATATTGGAAGAAAAATAAAAGATAGAGCATCGAGTGATGCAAAAAAT 342
IAM17upsAB7 AATGTTTTGGATGTTCTTGCAGAAGATGTAAAGAAACAGGCAGAAAAAGCTGCTGAAAGT 342
               *** * **       *         **
```

```
ISM51upsAG3 CATACAAATAAATTGGAAGGCACATTAAGAGAGGCAAGATTTCTTGATGGCTTGCATAAG 399
ISM49upsAF3 TATACAAGTCAATTGAGAGGCGATTTATCAAGCGCACGATTTCATGATGGCTTGCACAAG 399
ISM16upsAG6 CATAAAGATCAATTGATAGGCGTATTAAGAAATGCAAGATTTGCTGATCGATTGTATAAG 399
ISM11upsAH5 TATATAAATAGATTAAAAGGTAATTTATCAAATGCAATATTCATTGATGGATTGAGTAGA 399
ISM2upsAB12 TATGACGGGAAATTGAAAGGAAAATTATCAAATGCAAAATTTGCTGATGGCTTGTATAAG 399
ISM16upsAE10CATAGTGAAATATTGAAAGGCACATTATCAAAAGCACAATTTCTTGATGGCTTGCATAAG 399
ISM3upsAC6  TATAATAGTGAATTGATAGGCACATTATCAAACGCCCAATTTCATGATGGCTTGCGCAAG 399
ISM49upsAF8 TATGCAAGTCAATTGCGAGGCAATTTATCAAGAGCAACATTTTGTGATGTCTTTTGTGAT 408
IAM17upsAA3 CATGAAAGGCAATTGAAAGGAAAATTATCAAATGCAAAATTTGCTGATCGATTGTATAAG 399
IAM17upsAD3 CATGTGGATTCGTTGAAAGGGGATTTAACGAAAGCAGAATTTCGTGGTGGTCCTTCTACG 405
ISM49upsAG3 CATGTGGATTCGTTGAAAGGGGATTTAACGAAAGCAGAATTTCGTGGTGGTCCTTCTACG 405
ISM11upsAG9 TATGCAAGTTCTTTGAAAGGAGATTTGAAGAGAGCAAAATTTAATCATGATTTTTTTA-- 415
ISM3upsAB11 CATGTAAGTTCTTTGAAAGGAAATTTATTACAAGCACAATTTTATCA-TGCGTACTCTAT 401
ISM3upsAD6  CATAGTAGCTCTTTGAAAGGATATTTGTCACAAGCAAAATTTCATCAGCCCTTATTAAAG 402
ISM2upsAD10 CATGATAAATCTTTGCAAGGAAATTTTAAGAAAGCACAATTTCATCAGCGCTTATTAAGG 405
ISM16upsAE3 CATGGAAAAATTTTGAAAGGAAATTTGAGGCAAGCAAAATTTTATCA-TGATTTTTCAAA 398
ISM2upsAD3  TATAGAAGTTATTTGAAAGGAAATTTGGGGAATGCACAATTTTATCA-TGAGTACTCTAA 401
IAM17upsAB7 TATGAAAATGAATTGAAAGGAAAATTAGAAGAAGCATCATTTTGTGA-TGCTTATTGTGA 401
             **          **   ***    **        **   ***   *

ISM51upsAG3 GCAGCTCGTGGGGACATAAAATATGGTCCTCAAGATTCGTGTAGTCTTGATCACAAATTC 459
ISM49upsAF3 GCAGCTGATTTGGGGGTAAGACCTGGTCCTGGAGATTCGTGCAGTATTAATCACTTATTT 459
ISM16upsAG6 GAATCTGTTGGGGGGGTAAGAACTGGTCCTGCAAATTCATGCGACCTTGAACACAAATAC 459
ISM11upsAH5 GCACATAATTTAA---CAAGACGTGGTCCTGCAAATTCATGCGATCTTGAACACAGATTC 456
ISM2upsAB12 GAATCTGTTGGGGGGGTAAGATATGGTCCTGCAAATTCATGCGATCTTGAACACAGATTC 459
ISM16upsAE10GCAGCTCGTTTGAGGGTAACATATGTTCCTGAAGATTCGTGTAAGCTTAGTCACAAATTC 459
ISM3upsAC6  GCAGCTGGTTGGGGGGTAATACCTGGTCCTGCAAATTCATGCGATCTTGAACACAGATTC 459
ISM49upsAF8 TTTGTTGG--AATTCGCAAT-TATCCTTATACTGATCCATGTTATCTTGATCATAGATTT 465
IAM17upsAA3 GAATCTGGTGGTGATTTAAGGTCTGCTTATTCAGATGCTTGTTCACTTACATACAAATTT 459
IAM17upsAD3 CCAGTAAATAAGCA---TAATTACTATTATTCATATCCATGTAATTTAGATCATAAGGAA 462
ISM49upsAG3 CCAGTAAATAAGCA---TAATTACTATTATTCATATCCATGTAATTTAGATCATAAGGAA 462
ISM11upsAG9 -----AAATAA--A---AAGTTACATGCCTAGAAATCCATGTTATCTTGATTATGCTTTT 465
ISM3upsAB11 GTATAGAACTGTCC--CTGGA------------AATCCATGTAATCTTGATTATATATTT 447
ISM3upsAD6  GCAGCGAAATATGT--TTGGATA-GCTCCTAGTAATCCATGTTATCTTGATTATATATTT 459
ISM2upsAD10 GCAACGAAATATGT--TTGGACA-CCTCCTAGTAATCCTTGCTATTTTGATTTTAAGTTT 462
ISM16upsAE3 ACTTTATCCTAACT--ATAGA------------AGTCCCTGTGATCTTAATTTTTGGTTT 444
ISM2upsAD3  GTATAGAACTGTCC--CTGAA------------AGTCCATGTGATCTTAATTTTTGGTTT 447
IAM17upsAB7 ATGGATAGGTGTGT--CTAAATATGGTTCTACAGATCCATGTTATCTGGATCATATGAGG 459
                                            * * **     *

ISM51upsAG3 CATACTAAT-ATAAACAATGGTACTAATCATGGAAGAAATCCTTGCGATCTTAGAAATCA 518
ISM49upsAF3 CATACTAAT-ATAAACAATGGTACTAATCATGGAGGAAATCCTTGCGATCTTAGAAATCA 518
ISM16upsAG6 TATACTAAT-ATAGATATTGGATATCTACCTGCGAGGAATCCTTGTCATGGAAGAAAAGA 518
ISM11upsAH5 TATACTAAT-ATAAATAATGGATATCTACCTGCGAGGAATCCTTGCCATAATAGAAATCA 515
ISM2upsAB12 TATACTAAT-ATAAATAATGGATATCTACCTGCGAGGAATCCTTGTCATGGAAGAAAAGA 518
ISM16upsAE10CATACTAAT-ATAACAAAT------CAACATGAAAGGAATCCTTGTCATGGAAGAAAAGA 512
ISM3upsAC6  TATACTAAT-ATAAATAATGGATATCTACCTGCGAGGAATCCTTGCCATAATAGAAATCA 518
ISM49upsAF8 TATACAAAT-ATAAAAGTTAATTCTATAGAAGGAAGAAACCCTTGTAATGGTAGAGAAAA 524
IAM17upsAA3 CATACTAAT-ATAACAACTGATGGGGGGGGATGGAAGGCATCCTTGTCATGGTAGGGAAAA 518
IAM17upsAD3 CATACTAAT-TTACGGTATGATGATGTGAATTTGAGACATCCTTGCCATGGTAGAGAACA 521
ISM49upsAG3 CATACTAAT-TTACGGTATGATGATGTGAATTTGAGACATCCTTGCCATGGTAGAGAACA 521
ISM11upsAG9 CATTCTAATACTCCTGGAAATCGAAGAGAATTT-AGACATCCTTGTGCTGGTAGAAACAA 524
ISM3upsAB11 CATACTAAT-GTATGGCATCGTAACGCAGAAGATAGAAATCCTTGTCTTTTTAGTCGTGC 506
ISM3upsAD6  CATACTAAT-GTATGGAATGAGCGTGCACATGATAGAGATCCTTGTCTTTTTAGTCGTGC 518
ISM2upsAD10 CATACTAAT-GCTCCAAATGATCGTTCAAAGGATAGACATCCTTGTTATTTGAGAGATAT 521
ISM16upsAE3 CATACGAAT-GTTTGGAAGCGAACACCACGTGAAAGAGATCCTTGTTATCGTAGGCAACC 503
ISM2upsAD3  CATACTAAT-GTATGGCATGGTAAAGCAGAAGATAGAAATCCTTGTCTCTTTAGTGATAA 506
IAM17upsAB7 AATACAAAT-TTATTGAATAATAGAGTAAATGAGAGAAATCCTTGCCATGGCGGAAATCA 518
             ** * ***                          *  * *****        *
```

```
ISM51upsAG3 AAATCGTTTTGGTGAAAATGCCGAAGCGTATTGTAATAGTGATAAAATAAGAGGTAATGA 578
ISM49upsAF3 AAATCGTTTTTCTGAAAATGCCGAAGCGTATTGTAATAGTGATAAAATAAGAGATTAT-A 577
ISM16upsAG6 AAATCGTTTTTCTGAAAATGCCGAAGCGTATTGTAATAGTGATAAAATAAGAGATTATAG 578
ISM11upsAH5 AAATCGTTTTGACGAAAATGCCGAAGCGTATTGTAACAATGATAAAATAAGAGATAATGG 575
ISM2upsAB12 AAATCGTTTTGACGAAAATGCCGAAGCGTATTGTAATAGTGATAAAATAAGGGTTATTGG 578
ISM16upsAE10AAATCGTTTTGACGAAAATGCCGAAGCGTATTGTAATAGTGATAAAATAAGAGGTAATGA 572
ISM3upsAC6  AAATCGTTTTGACGAAAATGCCGAAGCGTATTGTAATAGTGATAAAATAAGAGGTAATGA 578
ISM49upsAF8 AAAACGTTTTGGTGAAAATGCCGAAGCGTATTGTAATAGTGATAAAATAAGAGGTAATGA 584
IAM17upsAA3 CAATCGTTTTTCTGAAAGTCAAGAATATGGATGTAGTAATGTATACATAAAAGGAAATGA 578
IAM17upsAD3 AAACCGATTTGATGAAGATGAAGAATCTGAATGTGGAA---ATAAAATACGTAATTATAA 578
ISM49upsAG3 AAACCGATTTGATGAAGATGAAGAATCTGAATGTGGAA---ATAAAATACGTAATTATAA 578
ISM11upsAG9 AACTCGTTTTTCAAACGAAAGTGAAGCAGAATGTGGTAGTGATAAAATAAGGGTTATTGG 584
ISM3upsAB11 AAAACGTTTTTCAAATGAAGGTGAAGCAGAATGTAATGGTGGTATAATAACTGGTAATAA 566
ISM3upsAD6  AAAACGTTTTTCAAATGAAGGTGAAGCAGAATGTAATGGTGGTATAATAACTGGTAATAA 578
ISM2upsAD10 AAATCGTTTTTCAGATAGAGGAGATGCAATATGTACTAATAATAAAATCAATTGTAATAA 581
ISM16upsAE3 AAAAAATAATCCGAATTTGGAAGGAGCAGTATGTACAAATAGTAAAATAAAAGGTAATGA 563
ISM2upsAD3  AAATCGTTTTTCAAATGGTGGTGGAGCAGAATGTGATAATAATAAAATAACTGGTAATGA 566
IAM17upsAB7 AAGACGTTTTGATGAAGATCAAGTATGTGAATGTGGTAAAAGTAGAATAAAAGGTAATGA 578
             *       *    *       *        ***             **         *

ISM51upsAG3 AAATAACAGAAATGATGGTACAGCATGTGTACCATTTAGAAGGCAAAATTTATGTGATAA 638
ISM49upsAF3 GGATA--AAAAGTGCTGGTACAGCATGTGCACCATTCCGAAGACAAAATTTGTGTGATAA 635
ISM16upsAG6 GATAA---AAAGTGCTGGTGGAGCATGTGTACCATTTAGAAGGCAAAATTTATGTGATAA 635
ISM11upsAH5 GGAAA---GAAGTGCTGGTGGAGCATGTGCACCATTCCGAAGGCAAAATATGTGTGATAA 632
ISM2upsAB12 AAATA---ACAGGAAGGATGGAGCATGTGCACCATTCCGAAGGCAAAATTTGTGTGATAA 635
ISM16upsAE10AAATAACAGAAATGATGGTACAGCATGTGCACCATTCAGAAGACAAAATTTGTGTGATAA 632
ISM3upsAC6  AAATA---ACAGTAATGCTGGAGCATGTGCGCCATTCCGAAGACAAAATTTGTGTGATAA 635
ISM49upsAF8 AAATA---ACAGTAATGCTGGAGCATGTGCACCATTTCGGAGGCAAAATTTGTGTGATAA 641
IAM17upsAA3 AAATA---ACAGTAATGGTACAGCATGCGTACCACCAAGAAGAAGACATATATGTGATCA 635
IAM17upsAD3 AAGAG---AAAATGATGCTATAGCCTGTGCGCCACCTAGAAGACGACATATGTGTGATAA 635
ISM49upsAG3 AAGAG---AAAATGATGCTATAGCCTGTGCGCCACCTAGAAGACGACATATGTGTGATAA 635
ISM11upsAG9 AAATA---ACAGAAATGATGGAGCATGTGCACCATATAGGAGAAGACATATATGTGACTT 641
ISM3upsAB11 AGG-----TGAATG----TGGGGCATGTGCACCGTATAGGAGAAGACATATATGTGACTA 617
ISM3upsAD6  AGG-----TGAATG----TGGGGCATGTGCACCGTATAGGAGAAGACATATATGTGACTA 629
ISM2upsAD10 TGA-----TGGTTG----CGGAGCCTGTGCTCCATATAGAAGAATACAGTTATGCGATTA 632
ISM16upsAE3 AAACAAAATAATTGACATTGGAGCGTGTGCCCCATATAGAAGACGAAATATCTGCGATTA 623
ISM2upsAD3  AGG-----TAAAAT----TGGAGCATGTGCTCCATATAGAAGGAGAGAATTGTGTGATTA 617
IAM17upsAB7 AAAT---AAAAATGATGGTGGATCCTGTGCCCCTCCAAGAAGAAAACATATATGTGATAA 635
                              * ** *  **     * **   * *  * ** **

ISM51upsAG3 AAATTTAGAATATTTAGATAACAAAAACACGAACACTACTGATGATTTATTGGGAAATGT 698
ISM49upsAF3 AAATTTAGAATATTTGATTAATGAAAACACAAATACTACTCACGATTTATTGGGAAATGT 695
ISM16upsAG6 AAATTTAGAATATTTGATTAATGAAAACACGAAAACTACTCATGATTTATTGGGAAATGT 695
ISM11upsAH5 AAATTTAGAATATTTGATTAATGAAAACACGTAAACTACCCATGATTTATTAGGAAATGT 692
ISM2upsAB12 AAATTTAGAATATTTGATTAATGAAAACACAGAAAATACGCATGATTTATTGGGAAATGT 695
ISM16upsAE10AAATTTAGAATATTTGATTAATGAAAACACAAATACTACGCATGATTTATTGGGAAATGT 692
ISM3upsAC6  AAATTTAGAATATTTGATCAATAAAAACACAGAAAATACTCATGACTTATTGGGAAATGT 695
ISM49upsAF8 AAACCTGGAAGCTATAAATGCAACTAATACAAAAAATACTCATGATTTATTGGGAAATGT 701
IAM17upsAA3 AAATTTAGAATTTTTAGATAATCCTCACACTGATGATACTGATGATTTGTTGGGAAATGT 695
IAM17upsAD3 AAACTTGGAAGCTCTAAATGATATAAATACCCAAAATATTCATGATTTGTTAGGAAATGT 695
ISM49upsAG3 AAACTTGGAAGCTCTAAATGATATAAATACCCAAAATATTCATGATTTATTGGGAAACGT 695
ISM11upsAG9 AAATTTGGAATACATAGATGTACATAATACAAAAAATTCTAACGACTTGTTAGGAAATAT 701
ISM3upsAB11 TAATTTGCACCATATAAACGAAAATAATATAAGGAATACTCATGATTTATTGGGGAATTT 677
ISM3upsAD6  TAATTTGCACCATATAAACGAAAATAATATAAGGAATACTCATGATTTATTGGGGAATTT 689
ISM2upsAD10 TAATTTAGAGCATATAAATGATAGTAATATTAATAGTACTGATGATTTATTGGGGAATCT 692
ISM16upsAE3 CAATTTAGAACATCTAAATGAAAGAAATGTTTTAAATACTCATGATTTATTGGGAAATGT 683
ISM2upsAD3  TAATTTGGAACATATAGATGTAAATAATGTGAAAAGTATTCATGATTTATTGGGGAATTT 677
IAM17upsAB7 AAACTTGGAAGCACTAAATGAAAGTAATACCAAAAATACTCATGATTTATTGGGAAATGT 695
             **  *  *     *           *         *    * ** ** ** ** **  *
```

```
ISM51upsAG3 GTTAGTTACAGCAAAATATGAAGGTGAATCTATTGTTAAAAATCATCCACATA------A 752
ISM49upsAF3 ACTAGTTACAGCAAAATATGAAGGTGAATCTATTGTTAGGAATCATCCAAATA------G 749
ISM16upsAG6 GTTAGTTACAGCAAAATATGAAGGTGAATCTATTGTTAGGAATCATCCAAATA------A 749
ISM11upsAH5 ACTAGTTACAGCAAAATACGAAGGTGAATCTATTGTTAAAAATCATCCAAACA------A 746
ISM2upsAB12 GTTAGTTACAGCAAAATATGAAGGTGAATCTATTGTTAAAAATCATCCACATA------A 749
ISM16upsAE10GTTAGTTACAGCAAAATATGAAGGTGAATCTATTGTTGCGAAGCATCCACATA------A 746
ISM3upsAC6  GTTAGTTACAGCAAAATATGAAGGCGAATCTATTGTTGCGAAGCATCCACATA------A 749
ISM49upsAF8 GTTAGTTACAGCAAAATATGAAGGTGATATTATTGTTAGTAATCATCCAAATA------C 755
IAM17upsAA3 GTTAGTTACAGCAAAATATGAAGGTCAATCTATTGTTGAAAAGCATCCAAATA------A 749
IAM17upsAD3 GTTAGTTACAGCAAAATACGAAGGTGAATCAATTGTTAATAATCATCCACATA------A 749
ISM49upsAG3 ACTAGTTACAGCAAAATACGAAGGTGAATCAATTGTTAATAATCATCCACATA------A 749
ISM11upsAG9 CTTGGTCACAGCAAAATATGAAGGTGAATCTATTGTTGAAAAGCATCCAAATA------A 755
ISM3upsAB11 GTTAGTTATGGCAAGGAGTGAAGGTGAATCTATTGTGAAAAGTCATGAATATACAGGTTA 737
ISM3upsAD6  GTTAGTTATGGCAAGGAGTGAAGGTGAATCTATTGTGAAAAGTCATGAATATACAGGTTA 749
ISM2upsAD10 ATTAGTTATGGCAAAAAGTGAGGGTGATTCGATTGTGAAAAGTCATGAAAATACAGGTAA 752
ISM16upsAE3 GTTAGTTATGGCAAAACGTGAAGGAGAATCTATTGTGAATAGTCAAGCAAATA------A 737
ISM2upsAD3  GTTAGTTATGGCAAGGAGTGAAGGTGAATCTATTGTGAATAGTCATAAAAATAC------ 731
IAM17upsAB7 ATTAGTTACAGCAAAATATGAAGGCGAATCTATTGTGAATAGTCATAAACATA------G 749
              * ** *  ****     ** **  *    *****    *  **   * * *

ISM51upsAG3 AGGAACTT-CAGAC-----GTATGTACTGCACTTGCACGCAGTTTTGCAGATATAGGTGA 806
ISM49upsAF3 AGGATCTT-CAGAA-----GTATGTATTGCCCTTGCACGAAGTTTTGCCGATATAGGTGA 803
ISM16upsAG6 AGATACTT-CAGAA-----GTATGTACTGCACTCTCACGAAGTTTTGCAGATATTGGAGA 803
ISM11upsAH5 AAAAACTT-CAGAC-----GTTTGTACTGCCCTTGCAAGAAGTTTTGCAGATATAGGTGA 800
ISM2upsAB12 AGGAACTT-CAGAC-----GTATGTACTGCACTTGCACGCAGTTTTGCCGATATAGGTGA 803
ISM16upsAE10AGACAATT-CACAA-----GTATGTACTTCTCTTGCACGAAGTTTTGCTGATATAGGTGA 800
ISM3upsAC6  AGACAATT-CACAA-----GTATGTATTGCCCTTGCAAGAAGTTTTGCAGATATAGGTGA 803
ISM49upsAF8 AAAAACGT-CAGAC-----GTATGTACTGCCCTTGCACGAAGTTTTGCAGATATAGGAGA 809
IAM17upsAA3 AGAAACAT-CAGAC-----GTTTGTACTGCCCTTGCACGAAGTTTTGCAGATATAGGAGA 803
IAM17upsAD3 AGGAACTT-CAGAC-----GTTTGTACTGCTCTTGCACGAAGTTTTGCAGATATAGGTGA 803
ISM49upsAG3 AGGAACTT-CAGAC-----GTTTGTACTGCTCTTGCACGAAGTCTTGCAGATATAGGTGA 803
ISM11upsAG9 AGAAACAT-CAGAC-----GTTTGTACTGCACTTGCAAGAAGTTTTGCAGATATAGGAGA 809
ISM3upsAB11 TGGTATATACAAATCAGGTATATGTACTTCTCTTGCTCGCAGTTTTGCAGATATAGGTGA 797
ISM3upsAD6  TGGTATATACAAATCAGGTATATGTACTTCTCTTGCTCGCAGTTTTGCAGATATAGGTGA 809
ISM2upsAD10 TGGTATTTATAAATCGGGGGTATGTACTTCTCTTGCTCGCAGTTTTGCAGATATAGGTGA 812
ISM16upsAE3 TGGAACGT-TAAAT-----GTATGTACTGCACTTGCACGAAGTTTTGCAGATATAGGTGA 791
ISM2upsAD3  TGGCATG-ATAAAC-----GTATGTGCTTCTCTTGCACGAAGTTTTGCTGATATAGGCGA 785
IAM17upsAB7 TGGAATGT-TAAAC-----GTATGTATTGCCCTTGCACGAAGTTTTGCAGATATAGGTGA 803
                    * *       * ***  * * **  *  * *** **** ***** ** **

ISM51upsAG3 TATTGTAAGAGGAAAAGATATGTTT---AAA---CGTAATGAAGAAGACGCAGTGCAGAA 860
ISM49upsAF3 TATTGTAAGAGGAAGGGATATGTTT---AAA---CGTAATGACAAAGATGCAGTGCAGAA 857
ISM16upsAG6 TATTGTAAGAGGAAAAGATATGTTT---AAA---CGTAATGAAGAAGACGCAGTGCAGAA 857
ISM11upsAH5 TATTGTAAGAGGAAGAGATATGTTT---AAA---TCTAATGAGGATGTCG------AAAA 848
ISM2upsAB12 TATTGTAAGAGGAAGAGATATGTTT---AAA---CCTAATGACAAAGATGCAGTGCGGCA 857
ISM16upsAE10TATTGTAAGAGGGAGAGATATGTTT---AAA---CCAAATGAAGAAGACGCAGTGCAGAA 854
ISM3upsAC6  TATTGTAAGAGGAAGGGATATGTTT---AAA---CGTAATAACCATGACAATGTAGAAAA 857
ISM49upsAF8 TATTGTAAGAGGAAGAGATATGTTT---AAA---TCTAAT------GAGAATGTAGAAAA 857
IAM17upsAA3 TATTGTAAGAGGAAGAGATATGTTT---AAA---TCTAAT------GACAAAGTAGAAAA 851
IAM17upsAD3 TATTATAAGAGGAATAGATATGTTT---AAA---CCAAATGTCCATGACAAAGTAGAAAA 857
ISM49upsAG3 TATTATAAGAGGAATAGATATGTTT---AAA---CCAAATGTCCATGACAAAGTAGAAAA 857
ISM11upsAG9 TATTATCAGAGGAAAAGATCTGTTTCTTGGT---AATAATGATAATGATAAAGTAAAAAA 866
ISM3upsAB11 TATTATTAGAGGAAAAGATCTTTTTCTGGGG---CATAAAGAACAAAAAAAAAAATTACA 854
ISM3upsAD6  TATTATTAGAGGAAAAGATCTTTTTCTGGGG---CATAAAGAACAAAAAAAAAAATTACA 866
ISM2upsAD10 TATAATCAGAGGAAAAGATCTTTTTCTGGGG---CATAATCATAAAAAAAAACCATTACT 869
ISM16upsAE3 TATTGTAAGAGGAACAGATCTTTTCCTTGGTGGTCCTAGTCAAGAGAAAAAAAAATTAGA 851
ISM2upsAD3  TATTGTAAGAGGAAAAGATTTGTATCTCGGT---GATAAAGGAGAAAAAAAAAAATTAGA 842
IAM17upsAB7 TATTGTAAGAGGAAAAGATCTTTTCCTTGGTGGACCTAAGCAAGAGAAAAAAGAATTAGA 863
            ***  * ***** *  *** * *             *
```

```
ISM51upsAG3 AGG---------------TCTAAGGGCAGTTTTCAAGAAAATACATAGTAG-------TT 898
ISM49upsAF3 AGG---------------TCTAAGGGCAGTGTTTAAGAAAATAAATGACAA-------CT 895
ISM16upsAG6 AGG---------------TCTGAGGGCGGTTTTTAAGAAAATAAATGACAA-------CT 895
ISM11upsAH5 AGG---------------ACTAAAAGTAGTTTTCCAGAAAATATATGATGA-------CT 886
ISM2upsAB12 TGG---------------TTTAAAGGTAGTTTTTAAGAAAATATATGATAAATTGTCACC 902
ISM16upsAE10AGG---------------TCTAAGAGAAGTTTTCAAGAAAATAAAAGACGA-------CT 892
ISM3upsAC6  CGG---------------TCTAAGAGAAGTTTTTAGGAAAATATATAATGG-------CT 895
ISM49upsAF8 GGG---------------ACTAAAAGCAGTTTTTAAGAAAATAAATAATGG-------AT 895
IAM17upsAA3 AGG---------------ACTACAAGTAGTTTTCGGGAAAATAAAAGACGA-------CT 889
IAM17upsAD3 GGG---------------TCTCCGAGAGGTTTTCAAGAAAATACATGACTTG------AA 896
ISM49upsAG3 GGG---------------TCTCCGAGAGGTTTTCAAGAAAATACATGATGGA------AT 896
ISM11upsAG9 AGAGAAACTACAAAATAATTTAAAAAGTATTTTTGCCAAAATATATAAGGAA------TT 920
ISM3upsAB11 GGAAAA------------TTTAGAAAAAATTTTTAACAAATTTAAGACAATAT----ATG 898
ISM3upsAD6  GGAAAA------------TTTAGAAAAAATTTTTAACAAATTTAAGACAATAT----ATG 910
ISM2upsAD10 GGATAA------------TTTAGAAAAGATTTTTAAAAAATTTAGGGAAAAAT----ATA 913
ISM16upsAE3 AGAAAA------------TCTGAAAAAAATATTTGAAAACATTAAGAATAAAA----ATA 895
ISM2upsAD3  GAAGAA------------TTTGAAAGATATTTTCAAGCAAATACATGAAAAAT----TGA 886
IAM17upsAB7 AGAAAA------------TCTAAAAAAGATATTTCATAATATTCAGAAAAGCG----ATA 907
                                *         * **     *  *

ISM51upsAG3 TGAATAAAAGCAAAATTAATGATTATGA------TGGTGATGGTCCAGAGTATTATAAAT 952
ISM49upsAF3 TAAATGAAAAAAGAATCACACATTATAA------TGATGGATCTGGAAATTATTATAAAT 949
ISM16upsAG6 TAAAGGAAAAAGAAATTA---GTGATTA------TGATAATGATCCAAATTATTATAAAT 946
ISM11upsAH5 TAAAGAAAAAAGAAATTA---GTGATTA------TGATAATGATCCAAATTATTATAAAT 937
ISM2upsAB12 TAAAGTACAAGAACATTACA-AAGATGT------TGATGGATCTGGAAATTATTATAAAT 955
ISM16upsAE10TGAACAAAAACGGAATTAATGATTATAA------TGATGAAAATGGAAATTATTATAAAT 946
ISM3upsAC6  TGAAGGACAAGGGAATTCATCATTATAAAGATGATGATATTTCTGGAAATTACTCCAAAT 955
ISM49upsAF8 TGAAGGACAAGGGAATTCGTTATTATTA---TGACAATGGATCTGGAAATTATTATAAAT 952
IAM17upsAA3 TGAAGAAACAAGGAATTATTCATTATGA------TCATGATGGTCCACACTATTACAAAT 943
IAM17upsAD3 TAAAAGCAAAATTAA----TGATTATGATGGT---GATGGTCC---AGAGTATTACAAAT 946
ISM49upsAG3 GGAAGGTGAGGTAAA----AAATTATTACGATCCTGATGGATCTGGAAATTATTATAAAT 952
ISM11upsAG9 GAAATTGGAAAATAACTC-TGACTACAAAGATGATGATGGATCTGGAAATTATTATAAAT 979
ISM3upsAB11 CAGGTCTTG-------AGGACGTTGCAAT-----TGATGATA------------------ 928
ISM3upsAD6  CAGGTCTTG-------AGGACGTTGCAAT-----TGATGATA------------------ 940
ISM2upsAD10 AAGACCTTA-------ATAACCTTCCAAT-----TGATGATA------------------ 943
ISM16upsAE3 CAAAACTTA-------GTACACTAACACT-----TGAAAAAG------------------ 925
ISM2upsAD3  CGGATCCAAGAGCAAAAGACCACTACAAAG---ATGAAAAAGACGGAAATTTTTTTCAAT 943
IAM17upsAB7 GTAGTCTTC-------AACGCCTTTCAAT-----AGAAAAAG------------------ 937

ISM51upsAG3 TAAGAGAAGCATGGTGGACAGCGAACAGAGACCAAGTATGGAGAGCTATAACATGTTATA 1012
ISM49upsAF3 TAAGGGAAGAGTGGTGGACTATAAATAGAGATCAAGTATGGAGAGCTATAACATGTTATA 1009
ISM16upsAG6 TAAGAGAAGATTGGTGGAATGTGAATAGAGAACAAGTATGGAGAGCTATAACATGTTATA 1006
ISM11upsAH5 TAAGGGAAGATTGGTGGACAGCGAACAGAGACCAAGTATGGAGAGCTATAACATGTTATA 997
ISM2upsAB12 TAAGAGAAGATTGGTGGGCACTTAACAGAGATCAAGTATGGAGAGCTATAACATGTTATA 1015
ISM16upsAE10TAAGGGAAGATTGGTGGAAAGCTAACCGAGACCAAGTATGGAAAGCCATAACGTGTGAAG 1006
ISM3upsAC6  TAAGGGAAGCTTGGTGGATAGCTAACAGAGATCAAGTCTGGAAAGCTATAACATGTGAAG 1015
ISM49upsAF8 TAAGAGAAGCATGGTGGAATGTGAATAGAAATAAAGTATGGGAAGCCATAACATGTAAAG 1012
IAM17upsAA3 TAAGAGAAGATTGGTGGACGGCTAACCGCGATCAAGTATGGAAAGCCATAACATGCAAAG 1003
IAM17upsAD3 TAAGAGAAGCATGGTGGACAGCGAACAGAGATCAAGTATGGAAAGCCATAACCTGTGGCG 1006
ISM49upsAG3 TAAGAGAAGCATGGTGGAATGTGAATAGAAATAAAGTATGGGAAGCCATAACCTGTGGCG 1012
ISM11upsAG9 TAAGAGAAGCTTGGTGGACATCGAACAGGGATCAAGTATGGAAAGCCATAACATGCAAAG 1039
ISM3upsAB11 TTAGAGAATACTGGTGGGCGCTTAATAGAGAAGACGTATGGAAAGCCATAACATGCAAAG 988
ISM3upsAD6  TTAGAGAATACTGGTGGGCGCTTAATAGAGAAGACGTATGGAAAGCCATAACATGCAAAG 1000
ISM2upsAD10 TAAGAGAATATTGGTGGGCACTTAATAGAAACGATGTATGGGAAGCATTGACATGCTCTG 1003
ISM16upsAE3 TTAGGGAATACTGGTGGGCACTTAATAGAAATGATGTATGGAAAGCATTAACATGTTCTG 985
ISM2upsAD3  TAAGGGAAGATTGGTGGACGGCTAACCGCGATCAAGTATGGAAAGCCATAACATGCAATG 1003
IAM17upsAB7 TTAGGGAATACTGGTGGGCAATTCATAGAAGAGAGGTATGGGATGCATTAACATGCAAGG 997
            * ** ***   ******       *  *     * ** ***   **  * ** **
```

```
ISM51upsAG3 TTCCGTATTATGTTAATTATTTTAAAAAAAAGTCAGGCGATACTATCGTTTTTACCAATG 1072
ISM49upsAF3 TTCCGTATTATGTTAATTATTTTAAAAATATATCGGAGAAATCTAGGAATTTTACCGATG 1069
ISM16upsAG6 TTCCGTATTATGTTAATTATTTTAAAAAAACGTCAGACAATACTATCGTTTTTACCAATG 1066
ISM11upsAH5 TTCCGTATTATGTTAATTATTTTAAAAAAAAGTCAGACAATACTATCGTTTTTACCAATG 1057
ISM2upsAB12 TTCCGTATTATGTTAATTATTTTAAAAAAACGTCAGACGATACTATCGTTTTTACCAATG 1075
ISM16upsAE10CTCCACAAAAAGTTGATTATTTTAGAAAAGGTTCAAATGGAGAAAGTATTTTTTCAAATA 1066
ISM3upsAC6  CTCCACAAAAAGTTGATTATTTTAGAAAAGGTTCAAATGGAGAAAGTATTTTTTCAAATA 1075
ISM49upsAF8 CTCCACAAAAAGCTAATTATTTTAGAAAAGGTTTAGATGGTAGTGATGTTTTTACAAGTC 1072
IAM17upsAA3 CTCCACAAAAGGTTGATTATTTTAGAAAAGGTTTAGATGGAAAAATAATTTTTTCAGATT 1063
IAM17upsAD3 CACTACCTAAATCTGCATATGTCTTGCAATCAGAAAATAATACACAATTACCTTCATATC 1066
ISM49upsAG3 CACTACCTAAATCTGCATATTTCATGCAATCAGAAGATAATAAACAATTATTTTCATATC 1072
ISM11upsAG9 CACCAGACAAAGCTAATTATTTCATATATAAATCGGACAAATTTCGTAAGTTTTCTAGTG 1099
ISM3upsAB11 CACCAACAGGTGCTGATTACTTCGTTTATAAATCAGGTAGTTTACTTAACTTTTCTAGT- 1047
ISM3upsAD6  CACCAACAGGTGCTGATTATTTCGTTTATAAATCAGGTAGTTTACTTAACTTTTCTAGT- 1059
ISM2upsAD10 CTCCAGGAGATGCTAAATATGTGAAATATTTTCCAAGTAATACAACGATTGTTTCATTT- 1062
ISM16upsAE3 CACCATATGAAGCTCAATATTTCATAAAATCAAGCGATAAAGAACACTCATTTTCAAGT- 1044
ISM2upsAD3  CACCATATAAAGCTTGGTATTTTATGCATTCAGAAGATAATAAACAATTATTTTCAGAT- 1062
IAM17upsAB7 CACCTACTGGTGCTGATTATTTCGTATATAAACCAGATAGGTTACGTAACTTTTCAAGTC 1057
              *          *   **  *     *                        * *   *

ISM51upsAG3 ACGGAAAATGTGGCCATTATGAAGGTGCTCCTCC---TACCAATTTAGATTACGTCCCTC 1129
ISM49upsAF3 ACGGAAAATGTGGCCATTATGAAGGTGCTCCTCC---TACCAATTTAGATTATGTTCCTC 1126
ISM16upsAG6 ACGGAAAATGTGGCCATTATGAAGGTGCTCCTCC---TACCAATTTAGATTATGTTCCTC 1123
ISM11upsAH5 ACGGAAAATGTGGCCATTATGAAGGTGCTCCTCC---TACCAATTTAGATTACGTCCCTC 1114
ISM2upsAB12 ACGGAAAATGTGGCCATTATGAAAATAATATTCT---TACAAATTTGGATTACGTCCCTC 1132
ISM16upsAE10GTGGAAAATGTGGTGGTAAGGAAGCACCCGTTCC---TACCTATTTAGATTACGTCCCTC 1123
ISM3upsAC6  GTGGAAAATGTGGTGGTAAGGAAGCACCCGTTCC---TACCAATTTAGATTACGTCCCTC 1132
ISM49upsAF8 AAGGATATTGTGGTCGTAAGGAACTAACCGTTCC---TACCTATTTAGATTACGTCCCTC 1129
IAM17upsAA3 ATGGACCATGTGGACGTTATGAAACAATCGTTCC---TACCTATTTAGATTATGTCCCTC 1120
IAM17upsAD3 TTAAATGCGGCCATAATAATAAGGATGATCCTCC---TACCAATTTAGATTATGTTCCTC 1123
ISM49upsAG3 CTAAATGCGGTCATAATAATAAGGATGATCCTCT---TACCAATTTAGATTATGTTCCTC 1129
ISM11upsAG9 ATAGGTGTGGACATAATGA---AGGTGATCCTCC---TACCAATTTAGATTATGTCCCTC 1153
ISM3upsAB11 --GATAGGTGTGGTCATAATAATAACGATGGTCCACTTACCAATTTAGATTACGTTCCTC 1105
ISM3upsAD6  --GATAGGTGTGGTCATAATAATAACGATGGTCCACTTACCAATTTAGATTACGTTCCTC 1117
ISM2upsAD10 --GATCAGTGTGGACATAATGATATGGATGTTCC---TACCAATTTAGATTACGTACCTC 1117
ISM16upsAE3 --GAATATTGTGGTCATTATAAAAACGGTGATCCACTTACCAATTTAGATTACGTGCCTC 1102
ISM2upsAD3  --TATAAATGCGGCCATTATGAA---GGTTCTCCTCTTACCAATTTGGATTATGTCCCTC 1117
IAM17upsAB7 ACGGAAAGTGCGGCCACAAGGAAGGAACTGTTCC---TACGAATCTAGATTACGTCCCTC 1114
                    *       *           **    ***  ** * ***** ** ****

ISM51upsAG3 AATTTTTACGTTGGTTCGAGGAATGGGCCGAAGATTTA 1167
ISM49upsAF3 AATTTTTACGTTGGTTCGAGGAATGGGCCGAAGATTTA 1164
ISM16upsAG6 AATTTTTGCGTTGGTTCGAGGAATGGGCCGAAGACTTA 1161
ISM11upsAH5 AATTTTTACGTTGGTTCGAGGAATGGGCCGAAGACTTA 1152
ISM2upsAB12 AATTTTTACGTTGGTTCGAGGAATGGGCCGAAGATTTA 1170
ISM16upsAE10AATTTTTGCGCTGGTTCGAGGAATGGGCCGAAGATTTA 1161
ISM3upsAC6  AATTTTTACGTTGGTTCGAGGAATGGGCCGAAGACTTA 1170
ISM49upsAF8 AATATTTACGTTGGTTCGAGGAATGGGCCGAAGACTTA 1167
IAM17upsAA3 AATTTTTAAGATGGTTCGAGGAATGGGCCGAAGACTTA 1158
IAM17upsAD3 AATATTTACGTTGGTTCGAGGAATGGGCCGAAGATTTA 1161
ISM49upsAG3 AATATTTACGTTGGTTCGAGGAATGGGCAGAAGACTTA 1167
ISM11upsAG9 AATTTTTACGTTGGTTCGAGGAATGGGCCGAAGATTTA 1191
ISM3upsAB11 AATTTTTGCGTTGGTTCGAGGAATGGGCCGAAGATTTA 1143
ISM3upsAD6  AATTTTTGCGTTGGTTCGAGGAATGGGCCGAAGATTTA 1155
ISM2upsAD10 AATTTTTAAGATGGTTCGAGGAATGGGCCGAAGATTTA 1155
ISM16upsAE3 AATTTTTAAGATGGTTCGAGGAATGGGCCGAAGACTTA 1140
ISM2upsAD3  AATTTTTACGTTGGTTCGAGGAATGGGCCGAAGATTTA 1155
IAM17upsAB7 AATTTTTAAGATGGTTCGAGGAATGGGCCGAAGATTTA 1152
              *** ***  * ***************** ***** ***
```

2.4 UpsB multiple sequence alignments of predominant sequences from two isolates

```
ISM4917DBLF3 ATGTAATTGTTGTTTTTTTTTTTGTTAGAATATTTAAATTTATTATAAACCTATTAATAT 60
ISM4917DBLE3 ATGTAATTGTTGTTTTTTTTTTTGTTAGAATATTTAAATTTATTATAAAACTA--AATAT 58
IAM1817DBLF5 ATGTAATTGTTGTTTTTTTTTTTGTTAGAATATTTAAATTTATTATAAACCTAT--TTGA 58
IAM1817DBLE12ATGTAATTGTTGTTTTTTTTTTTGTTAGAATATTTAAATTTATTATAAACCTATTAATAT 60
IAM1817DBLG11ATGTAATTGTTGTTTTTTTTTTTGTTAGAATATTTAAATTTATTATAAACCTATTAATAT 60
ISM4917DBLH12ATGTAATTGTTGTTTTTTTTTTTGTTAGAATATTTAAATTTATTATAAAAGTATTAATAT 60
             ********************************************  **    *

ISM4917DBLF3 ATATTTTTTTTAAAA-ATATATATATA----------AAACTAATTATTATTATTATATA 109
ISM4917DBLE3 ATAATTTTTTTT-----TTCTTTTTTA----------AAAATATATAT-------ATAAA 96
IAM1817DBLF5 ATATTTGATTAGA-----AAATATATA----------AAACTAATAATTAT---TATATA 100
IAM1817DBLE12AAATTTTTTTTAAAA-ATATATATATA----------AAACTAATAATTATTATTATATA 109
IAM1817DBLG11ATATTTTTTTTTAAAAATATATATATA----------AAACTAATAATTATTATTATATA 110
ISM4917DBLH12ATATTTTTTTAAAAATATATATATATATATATATAAAACTAATAATTATTATTATATA 120
             * * **  **          * * **          *** **   **       *** *

ISM4917DBLF3 CT----TATTACATATTATTTATTAATATATAT-ATATTATATATATTATTATAATATTAC 164
ISM4917DBLE3 AC----TA--AAATAATAATTACATACATATAC-ATATTA------------AATATTAT 137
IAM1817DBLF5 CA----TATTAAATATTATTTATTAATATATAC-ATATATATATAATATTATAATATTAC 155
IAM1817DBLE12CA----TATTAAATATTATGTATTAATATATATTATATATA----ATATTATAATATTAC 161
IAM1817DBLG11CA----TATTAAATATTATTTATTAATATATATAATATATA-----TATTATAATATTAC 161
ISM4917DBLH12CAAACATATGAAATATTATTTATTAATATATAT--TATATA-----TATTATAATATTAC 173
               **  * *** **  **   * *****   ***             *******

ISM4917DBLF3 TACTATTATAATTACTATATATATATATAAATATAAT--------ACTTATATATATATA 216
ISM4917DBLE3 T--TATTA----------ATATATATATAAATATA-------------TATATATATATA 172
IAM1817DBLF5 AACTATTATAATTACTATATATAAATATATATATAA------------TACTTATATATA 203
IAM1817DBLE12AACTATTATAATTAATATATATATATAAATATATATACAATACTTATATATATATATATA 221
IAM1817DBLG11TACTATTATAATTAATATATATATATAAACATATATATA------ATACTTATATATATA 215
ISM4917DBLH12TACTATTATAATTCCTATATATATATAAATATATATATA------ATACTTATATATATA 227
               *****          ***** *** * *****                ********

ISM4917DBLF3 TATTCCAACACAATACTATTATTATTATTCTACCCTATCACTATGCTCCCATAACATA-C 275
ISM4917DBLE3 CATCCAAA-------------------------TATTACAATACTCCCATAACATA-C 204
IAM1817DBLF5 TATTCCAAC--AAAAACAATATTATTATTCTACCATATCACTATACTCCCATAACATAAC 261
IAM1817DBLE12TATTCCAACACAATACTATTATTATTATTCTACCATATCACTATACTCCCATAACATA-C 280
IAM1817DBLG11T--TTCAACA-AATA-TAATATCATTATTCTACCATATCACTATACTCCCATAACATA-C 270
ISM4917DBLH12TATTCCAACACAATACTATTATTATTATTCTACCATATTACAATACTCCCATAACATA-C 286
                 **                          *** ** ** ************ *

ISM4917DBLF3 ATA-CATATATACATACATACCCCCACGTACGTACCAAAACACCACCAAACCATGTATGC 334
ISM4917DBLE3 ATA-CATA--------CATACCCCCACGTACGTACCAAAACACCACCAAACCATGTATGC 255
IAM1817DBLF5 ATAACATACATACATACATACCCCCACGTACGTACCAAAACACCACCAAACCATGTATGC 321
IAM1817DBLE12ATA-CATA--------TATACCCCCACGTACGTACCAAAACACCACCAAACCATGTATGC 331
IAM1817DBLG11GCA--ATA-------------CGCCAC-----CACCACCGCCCACACGAACCATGTATGC 310
ISM4917DBLH12GCA--ATA-------------CGCCAC-----CACCACCGCCCACACGAACCAAGCAAAC 326
               *  ***             * ****      ****   * *   * ***** * *  *

ISM4917DBLF3 CACGATATAAACCACGTA----TG-CTTGACATAATGTAGT------------CCCGAAA 377
ISM4917DBLE3 CACGATATAAACCACGTAC-CACG-TATGACATAATGTAGT------------GGTGG-- 299
IAM1817DBLF5 CACGATATAAACCACGTATGTATG-TATGACATAATGTAGT------------GGTGGAG 368
IAM1817DBLE12CACGATATAAACCACGTATGTATG-TATGACATAATGTAGTCACGAACAAAATGGTGAGG 390
IAM1817DBLG11CACGATATAAACCACGTATGTATG-CATGACATCATGTTGT------------CGCAAC- 356
ISM4917DBLH12C----CACAAAT---ATATGTATGACATGACATAATGTAGT------------CATGAA- 366
             *     * ***     **     *   ****** **** **

ISM4917DBLF3 CAAAAGAATCACAAAAATGGCGTCGCCCATGAGGC-----CAAGTAGTAGGGGTGGTGGT 432
ISM4917DBLE3 ---------TGTTAAAATGGCGGC-----TG---------CAGGTGGGGGTGGTAAGGAT 336
IAM1817DBLF5 T-------TAACAAAAATGGCGCC---------------CAAGAAGCCA-ACTGCGGAG 404
IAM1817DBLE12CCAGGAGCTGGTGGTGGTGGGAGTGGTCGTGGTGGGGGTGCTGCTGGTAGTAGTAAAGAG 450
IAM1817DBLG11--------CATGGAGAGTGGTAAGGG--------------CGGT-------GGCGGTGGT 387
ISM4917DBLH12--------TAACCAAAATGGTGAAGC--------------AAGTTAAAACTGGTGGCGTT 404
                           ***                                      *
```

```
ISM4917DBLF3 ACGCAGGAGGATCCTATT------GATAAAACAAGTGCCAAACATTTATTGGATAGCATA 486
ISM4917DBLE3 AAGTATGATGATGCCAAG------GAT---------GCAAAGGACCTTTTGGATAAAATT 381
IAM1817DBLF5 GACTATAATAAAGTCAAC------AAT---------GCGAAGGAGCTTTTTGATATGATT 449
IAM1817DBLE12AAGGAGGATGAACCTGACTATACTAATGTCAAGGATGCTAAGGAACTTTTGGATAAAATT 510
IAM1817DBLG11ACTAAGAAGGG---T---------------------GCGAAAGAAGTATTGGATGAATTT 423
ISM4917DBLH12ATTGAGGATGCAACT---------------------GCTAAACATATATTTGATGGGATT 443
                 *  *                            ** **  *  * ** ***    *

ISM4917DBLF3 GGGAAAATAGTGCATAA-------AAAAGCGCA---TCGTGACGCTCAAAAATATTATAC 536
ISM4917DBLE3 GGAGAAGATATATACAA-------AATAGCAAA---TAAGGCTGCTCTAAAGTACGAAAA 431
IAM1817DBLF5 GGAGAAACTGTACAGAA-------AAAAGTGCA---TGGTGCTACTAAAAAATATTATAC 499
IAM1817DBLE12GGGCAACAAGTGTACAAT----GAAAAAGTGAAAAATGGTGAAGCTCAAACATATGATAG 566
IAM1817DBLG11GGGCAACAAGTGTACAAAGAAGTGAAAAATG-A---TGCTGATGCTGAAAAATATAAAGA 479
ISM4917DBLH12GG--AAGGACTATACAA------CAACAAGTGCA---TAGTGCTTCTAAAATATATACTAG 493
             **  **    *  * **       *  *    *   *   *   **  **  **

ISM4917DBLF3 TCATTTACATGGAAGTTTGAAGGACGCAAAATTTGAGAAACTTCCAAACGGTCAACAAAC 596
ISM4917DBLE3 TGAATTGCATGGACGTTTGTCAGAAGCAAGATTTCAGGAAGATCCAAAGAAACAACAAAC 491
IAM1817DBLF5 TGAGTTACATGGAGATTTGTCACAAGCAACATTTGAGGGGA----AAAAAATTAGTGGAT 555
IAM1817DBLE12TTACTTGAAAGGAGATTTGAACAAAGCAAATGGTTATAGTT-----CGGAAACAGCTGGC 621
IAM1817DBLG11GGCGTTGAAAGGAAATTTGCAGGAAGCAAAA---------GGTATGGGGGAAAGAGCTGA 530
ISM4917DBLH12TGAATTGCATGGAGATTTAACAAAAGCAAAATTTCGACATGGTATATCGTGGGAAGAAGC 553
                 **  * ***  ***     * ****

ISM4917DBLF3 ACCACCAAATCCATGCGATCTTGATTATCAATGGCA---TACTAATGCTACTAACGTTAG 653
ISM4917DBLE3 CCCAGGAAATCCATGCAAACTTAATCATGAATATCA---TACTAATGCTACTAACGGTAA 548
IAM1817DBLF5 C-----ACCTCCATGTAATCTTGATTATACTAAACA---TACTAATGTTACCATTGGTGG 607
IAM1817DBLE12ACCATTAA-TCCATGCACACTTGTAGAGGAATATCG---TAGTAAGGCTAATGGTGATGG 677
IAM1817DBLG11CACCGATGATCCATGCGATTTTGATTATACTAAGGAACTCATTGGTGCTGCTGGCGGTGA 590
ISM4917DBLH12AAACAAAAATGTATGTGGTCTAAAACATTCATATGA---TACTAATGTTAGGTGGGGTGA 610
                   *  ***     *     *              * *   * *      * *

ISM4917DBLF3 AAGT------TATCCATGTAGA------GTTGGAAAAGA------AGAACGTTTTTCTCA 695
ISM4917DBLE3 AAGT------TATCCATGTAGA------ACTGGAAAAGA------AGAACGTTTTTCTCA 590
IAM1817DBLF5 GGGTAGGGAGTATCCTTGTAGA------AATGGAACAAA------AGAGCGTTTTCCCGA 655
IAM1817DBLE12TGAAAGG---TATCCGTGCACAGAGTTAAGTGGAAAAGAT---GTAGAACGTTTTTCGGA 731
IAM1817DBLG11ACGG------CATCCGTGCAAAAATTTAAAAGAAAATACAAATGAAAAACGTTTTTCCAA 644
ISM4917DBLH12AAGT------TATCCTTG-------TGCAAAGAGATCA-----GAAAAACGTTTTTCTGA 652
                  **** **           *  *         * * ****** *  *

ISM4917DBLF3 AGTACATGGCGGTGAATGTGATGATAAGAAAATAGAAGGTAATGGTCGTAATAATGGT-G 754
ISM4917DBLE3 AGTACATGGCGCTGAATGTGATAATAAAAAAATAAAAGATAATGAT---AGTAATGGT-G 646
IAM1817DBLF5 TACAGAAGGAGCACAATGTGATAAGAAAAAAATAAAAGATAGTAAAAACAGTAGCGAAAG 715
IAM1817DBLE12TACACTTGGTGGCCAATGCACTGATCAACAAATAGAAGGTAATGATCGTAAGAATGGT-G 790
IAM1817DBLG11TACACTTGGTGGTCAGTGTACTAAGGAAAAAATAAGTGGTAGTACA---AATACATGT-G 700
ISM4917DBLH12TGAAGGTGGCGGTGAATGTGATGATAGAAAAATAAAAGATAGTAAA---AACAATGGT-G 708
               *   ** *   * **   * *     *****   * ** *      *  *      *

ISM4917DBLF3 GAGCCTGCGCTCCCTTCAGAAGATTACATTTATGCGTTAGAAATTTGGAAAATATCAATA 814
ISM4917DBLE3 GAGCATGTGCTCCATATAGACGATTACATTTATGCGTTAGAAATTTAGAAAATATAAGTG 706
IAM1817DBLF5 GAGCATGTGCTCCTTTTAGACGATTACATCTATGCGACAAAAATCTGGAAAATATCAGTG 775
IAM1817DBLE12GCGCCTGTGCTCCATATAGACGATTATATCTATGTGATTATAATTTGGAAAAAATACCAA 850
IAM1817DBLG11GTGCTTGCGCTCCATACCGACGTCTACATCTATGTCATCATAATTTGGAAACTATA---G 757
ISM4917DBLH12GAGCCTGTGCTCCATATAGACGGTTACATTTATGTGATTATAATTTGGAAAAAATAACAG 768
             * ** ** ***** *   ** *  ** ** ****       *** * ****  **

ISM4917DBLF3 TTAATAATAAATACATTAATAAAGATACATTGTTGGCGGATGTATGTCTTGCAGCCAAAT 874
ISM4917DBLE3 CATTGCATAAG---ATTAATAAAGATACATTATTGGCGGATGTATGTCTTGCAGCCCTAC 763
IAM1817DBLF5 ATTTTAACAAT---ATTAATAATGATACATTATTGGTAGATGTGTGCCAGGCAGCCAAAC 832
IAM1817DBLE12CGTCGAAGACG------TCTACTGACACGTTGTTGGCAGAAGTGTGTATGGCAGCCAAAT 904
IAM1817DBLG11ACACAAAGTCG---ACGACG---CATGATTTATTGGCAGAGGTGTGTTATGCAGCAATAC 811
ISM4917DBLH12ATACACATACA---ACTACTACTCATAATTTATTGGTAGATGTGTGTCTTGCAGCAAAAT 825
                 *                 *    ** ****  ** ** **    *****   *

ISM4917DBLF3 ATGAAGGGGACTTAATAAAAACACATTATACACCATATCAA-CACAAA-----TATCGTG 928
ISM4917DBLE3 ATGAAGGACAATCAATAACACAAGATTATCCAAAATATCAAGCACAATATGCTTCTTCTG 823
IAM1817DBLF5 ATGAAGGACAATCAATAATACAAGATTATCCAAAATATCAAGCAACATATGCTTCTTCTG 892
IAM1817DBLE12ATGAAGGGGGCTTAATAAAAACACATTATACACGATATCAACAAATAT------ATGGTA 958
IAM1817DBLG11ATGAGGGAGAGTCACTACGAGGTCAACATGGAGAACATAAAGGAAC------------TA 859
ISM4917DBLH12ATGAAGGACAATCAATATCAGTTCAACATGGAAAATATCATACAGA------------TA 873
             **** **    * * **  *    *  **  *  * ** *   *             *
```

```
ISM4917DBLF3 ATTCTCCTTCTCAAATATGTACTATGTTAGCACGAAGTTTTGCAGATATAGGTGACATCG 988
ISM4917DBLE3 CTTCTCGTTCTCAAATATGTACTATGTTAGCACGCAGTTTTGCTGATATTGGAGACATTA 883
IAM1817DBLF5 GTTCTTCTTCTCAAATGTGTACTATGTTAGCACGAAGTTTTGCAGATATAGGAGACATTA 952
IAM1817DBLE12ATTCTCCTTCCCAATTATGTACGGTGTTGGCACGAAGTTTTGCAGATATAGGTGATATTA 1018
IAM1817DBLG11ATAATGAATCTCAATTATGTACTGTATTGGCACGCAGTTTTGCAGATATAGGTGATATCG 919
ISM4917DBLH12GTTCTGGTTCTACAATATGTACGGTATTGGCGCGAAGTTTTGCAGATATTGGAGACATTA 933
              *  *   **   * * *****  * ** ** ** ******** ***** ** ** **

ISM4917DBLF3 TACGCGGCAAAGATCTATATAGTGGTAAT----AAGAAA---------------AA---- 1025
ISM4917DBLE3 TCAGAGGAAAAGATCTGTATCGTGGTAAT----AATGGA---------------AA---- 920
IAM1817DBLF5 TACGAGGAAGAGATCTTTATATACGTAAT----AAGAAA---------------AA---- 989
IAM1817DBLE12TACGAGGAAAAGATCTGTATCTCGGTAATCCGCAAGAAAGT--------GCACAAA---- 1066
IAM1817DBLG11TACGTGGCAAAGATCTTTTCCTCGGTAATGATGAAGAAAAA--------AAACAAA---- 967
ISM4917DBLH12TTCGAGGAAAAGATCTATATCTTGGAAAAAAAAAAAAAAAAGCAAAATGAAACAAAAACA 993
              *  * ** * ****** *      * **     **   *               **

ISM4917DBLF3 ----------GGAAAA-ATTAGAACAGAATTTACAAAAAATTTTCAAAGAAATATATGAC 1074
ISM4917DBLE3 ----------GGATAA-ATTACAAGAAAATTTAAAAAAAATTTTCGAGAAAATATATGAC 969
IAM1817DBLF5 ----------AGATAA-ATTAGAAGATAATTTACAAAAAATTTTCAAGATGATTCAGGGA 1038
IAM1817DBLE12----------GAATAATATTAGAAAAGAATTTAAAAAAATATTTCCAACAAATACATAAT 1116
IAM1817DBLG11----------GAAAAAAATTACAACAACATTTGAAAGAAATTTTCGGGAAAATACATAAG 1017
ISM4917DBLH12GAAACAGAAAGAGAAAAATTAGAACAGAAATTGAAAGAAGTTTTCGCGAAAATACATAGT 1053
                         ** **** ** *  * **  ** **  ****      **  *

ISM4917DBLF3 AA------A---TTGGAT------AATAGCA------TAAAATCAAACTATAATGATGC- 1112
ISM4917DBLE3 AA------A---TTGGATGGGAAGAAAGGCG------CAAAAGACTACTACAAAGATGAA 1014
IAM1817DBLF5 AA------A---A-----------------A------TCAATCTAAACTAAGTCAACTT- 1065
IAM1817DBLE12GACGTGACG---TCTAGCGGGAGGAATGGTG------TAAAAGACCGCTACAAAGATAC- 1166
IAM1817DBLG11GAAGTGACG---ACAAATGGGAAGAATGTGAAGACGCTACAAGCTCGCTACGAAGGTGAT 1074
ISM4917DBLH12GAAGTGACGTCGACGAGCGGCAATAATAAGGAGGTGCTAAAAGCACGCTACGATGGTGAC 1113
              *                                        *      ***

ISM4917DBLF3 -----TCCATATTATTATCAATTACGTGAAGATTGGTGGAATAATAATAGAATAATGGTA 1167
ISM4917DBLE3 AATGGTGGAAATTATTATCAATTACGAGAAGATTGGTGGAACGCTAATAGACAAGAAGTA 1074
IAM1817DBLF5 --------ACACTAG-ATCAGGTAAGAGAATACTGGTGGGATGCAAATCGGCACACCGTG 1116
IAM1817DBLE12--TGACAAAAATTTTTTTCAATTACGAGAAGATTGGTGGTATGCTAATAGAGAAACAGTA 1224
IAM1817DBLG11AAAAAAAATTATTTTTTTCAATTAAGAGAAGATTGGTGGACAGCGAATCGAGAAACAGTA 1134
ISM4917DBLH12GGTGATAATTATTAT---CAATTAAGGGAAGATTGGTGGGATGCTAATAGACTTGATGTA 1170
                    * *     **  ** * *** * ******      *** *       **

ISM4917DBLF3 TGGTATGCTATGACGTGTGGTGAACCAGAAAAGGCTGAATATTTTAGAACAGCATGTTCT 1227
ISM4917DBLE3 TGGTACGCGATAACATGCGGCG---CTGGGGGTTATTCATATTTTCGACAAACATGTGGT 1131
IAM1817DBLF5 TGGAAAGCTATCACATGCAACG---CTGGAAGTTATAAATATTCTCGACCAACATGT--- 1170
IAM1817DBLE12TGGAAAGCCATCACATGCAACG---CTCAGGGTTTTGACTATTTTCGACAAACATGTGGT 1281
IAM1817DBLG11TGGAAAGCATTAACATGCAACGCTGGGAATG---CTAAATATTTTCGACCAACATGTGGT 1191
ISM4917DBLH12TGGAAAGCTATCACGTGCGGTGCACCACATGGTGCTCAATATTTTCGACAAACATGTAAT 1230
              *** * **  * ** **    *             *   **** * **  * *****

ISM4917DBLF3 ---GGTGG---AACAACTCCAACT------AATAAGAAATGCCGATGTGA---------- 1265
ISM4917DBLE3 ---GGAGG---AAAAACTGCGACT------GAAGGTAAATGCCGATGTCC---------- 1169
IAM1817DBLF5 ---AGTGA---ACAACCTTTGAGT------CAGGATAAGTGCCAATGTAT---------- 1208
IAM1817DBLE12GATGATGA---AAAAACTGCAACTCGGGTTAAAGACAAATGCCGGTGTGA---------- 1328
IAM1817DBLG11GGTGGTGATGAAAAAACTGGAATTCTGACTCGTAGTCAATGCCGGTGTGACGACAAGCCA 1251
ISM4917DBLH12GATGATGG------GACTTCTTCTCGTGCTATTCACCAATGCCGTTGTCAGAAGAAAGAC 1284
                  *         **     *             * ****  ***

ISM4917DBLF3 --------CAACGTAAGTATT------GTCCCCACCTATTTCGACTATGTGCCGCAGTTT 1311
ISM4917DBLE3 --------TAGTT---ATAAG------GTCCCTACATATTTTGACTATGTGCCACAATAT 1212
IAM1817DBLF5 --------CAATG---GCGGT------GTTCCCACTAATTTTGACTACGTGCCACAGTAT 1251
IAM1817DBLE12--------CGGCG---ACCAG------GTCCCCACATATTTTGACTACGTCCCTCAATTT 1371
IAM1817DBLG11AAGGCTGGCGACGGAGATGTAAATATTGTCCCCACATATTTTGACTATGTGCCACAGTAT 1311
ISM4917DBLH12GGCACACACGACAGCGACCAG------GTCCCCACATATTTTGACTATGTGCCACAGTAT 1338
                                       ** ** **  **** ***** ** ** ** * *

ISM4917DBLF3 CTTCGCTGGTTCGAGGAATGGGCCGAAGACTTA 1344
ISM4917DBLE3 TTGAGATGGTTCGAGGAATGGGCCGAAGACTTA 1245
IAM1817DBLF5 CTTCGCTGGTTCGAGGAATGGGCCGAAGATTTA 1284
IAM1817DBLE12TTAAGATGGTTCGAGGAATGGGCCGAAGACTTA 1404
IAM1817DBLG11CTTCGCTGGTTCGAGGAATGGGCCGAAGATTTA 1344
ISM4917DBLH12CTTCGCTGGTTCGAGGAATGGGCAGAAGATTTA 1371
              *  * ***************** ***** ***
```

2.5 upsC multiple sequence alignment of predominant sequences from 2 isolates

```
ISM335B1A2  CACATATAGTACGACTAAGAAACAAAATAATATCATAACAAACATAGTGACTACCGTTAC 60
ISM335B1B6  CACATATAATACGACTAAGAAACAAAATAATATCATAACAAACATAGTGACTACCGTTAC 60
ISM335B1B7  CACATATAGTACGACTAAGAAACAAAATAATATCATAACAAACATAGTGACTACCGTTAC 60
ISM25B1A11  CACATATAGTACGACTAAGAAACAAAATAACATCACAACAAACATAGTGACTACCATTAC 60
ISM25B1C9   CACATATAGTACGACTAAGAAACAAAATAACATCACAACAAACATAGTGACTACCGTTAC 60
ISM25B1C4   CACATATAGTACGACTAAGAAACAAAATAACATCATAACAAACATAGTGACTACCATTAA 60
            ******** ********************** **** ******************** ***

ISM335B1A2  AT-GATATTACCACATAATTCATACCATTATATAATATTACTACATGGTAATGATAACCA 119
ISM335B1B6  AT-GATATTACCACATAATTCATACCATTATATAATATTACTACATGGTAATGATAACCA 119
ISM335B1B7  AT-GATATTACCACATAATTCATACCATTATATAATATTACTACATGGTAATGATAACCA 119
ISM25B1A11  ATAGATATTACCACATAATATAAAGCATTAAATAATATTATTGCATGTTAGTGATAACTA 120
ISM25B1C9   AT-GATATTACCACATAATTCATACCACTATATAATATTACTACATGATAGTGATAACCA 119
ISM25B1C4   ATAGGTATTACCACACAATTCATACCACTATATAATATTACTACATGATAGTGGTAACTA 120
            ** * ********** ***  * * ** ** ********* * **** ** ** **** *

ISM335B1A2  CTATATCATATACACCACTATATAGTAATAGTAGCGGAGATATTATGTGCACAAATATAT 179
ISM335B1B6  CTATATCATATACACCACTATATAGTAATAGTAGCGGAGATATTATGTGCACAAATATAT 179
ISM335B1B7  CTATATCATATACACCACTATATAGTAATAGTAGCGGAGATATTATGTGCACAAATATAT 179
ISM25B1A11  CTATATCATATACACCACTATATAGTAATAGTAGCGGTGGTAATATGTACACGTATATAT 180
ISM25B1C9   TTATATCATATACACCACTATATAGTAATAGTAGCGAAGATATTATGTGCACAAATATAT 179
ISM25B1C4   CTATATCATATACACCATTATATAGTAATAGTAGCGGCGGTATCATGCACACGTATATAT 180
             ***************** ******************  * **  ***  ***  ******

ISM335B1A2  TATAATAGTGGTAGCCACAACCACGACATCATGGAAATATAGATTTTCATTCATATCTTC 239
ISM335B1B6  TATAATAGTGGTAGCCACAACCACGACATCATGGAAATATAGATTTTCATTCATATCTTC 239
ISM335B1B7  TATAATAGTGGTAGCCACAACCACGACATCATGGAAATATAGATTTTCATTCATATCTTC 239
ISM25B1A11  TGTAATAGTGGTAGCGACAATCACGATATCATGGTAATGTAGATTTTTATTCATATCTTC 240
ISM25B1C9   TATAATAGTGGTAGCAACAACCACGGTATCATGGTAATGTAGATTTTCATTCATATCTTC 239
ISM25B1C4   TGTAACAGTGGTAGCTACAATCACTGCATCATGGTAATATAGATTTTCGTTTATATCTTC 240
            * *** ********* **** ***   ******* *** ********  ** ********

ISM335B1A2  CTTATCGTTAGTTTTCCATACACTATTAATATGTATTTATGTTATAATGGTAGACTATGT 299
ISM335B1B6  CTTATCGTTAGTTTTCCATACACTATTAATATGTATTTATGTTATAATGGTAGACTATGT 299
ISM335B1B7  CTTATCGTTAGTTTTCCATACACTATTAATATGTATTTATGTTATAATGGTAGACTATGT 299
ISM25B1A11  CTTATCGTTTGTTGTCCATACACTTTTAATATGTATTTATGTTATAATGGTAAACTATGT 300
ISM25B1C9   CTTATCGTTAGTTTTCCATACCCTATTAATATGTATTTATGTTATAATGGTAGAATATGT 299
ISM25B1C4   TTTATTGTTTGTTGTCCATACACTATTAATATGTATTTATGTTATAATGGTAAACTATGT 300
             **** *** *** ******* ** ************************** * *****

ISM335B1A2  TAACAATGTATGAATGACTATCGTAAATTAATAATAGATACATGAAAAC------TGTGT 353
ISM335B1B6  TAACAATGTATGAATGACTATCGTAAATTAATAATAGATACATGAAAAC------TGTGT 353
ISM335B1B7  TAACAATGTATGAATGACTATCGTAAATTAATAATAGATACATGAAAAC------TGTGT 353
ISM25B1A11  TAACAATGTATGAATGATCATCGTAGATTAATAATAAATTCATGAAAACAATGTGTATGT 360
ISM25B1C9   TAACAATGTATGAATGACCATCGTAAATTAATAATAGATGCATGAAAACAATGTGTTTTT 359
ISM25B1C4   TAACAATGTATGAATGATCATCGTAGATTAATAATAGATGCATGGAAACCGTGTATATGT 360
            ****************  ****** ********** ** **** ****      * * *

ISM335B1A2  ATATGTATGTGTGC--------ATTTACAACATAATGTAGTCGGGAAGAAGAATACAAAA 405
ISM335B1B6  ATATGTATGTGTGC--------ATTTGCAACATAATGTAGTCGGGAAGAAGAATACAAAA 405
ISM335B1B7  ATATGTATGTGTGC--------ATTTACAACATAATGTAGTCGGAAAGAAGAATACAAAA 405
ISM25B1A11  ATATGTATGTGTGT--------ATTTACGACATAATGTAGTCGTG--GAAGCATACAAAA 410
ISM25B1C9   ATATGTATGTGTGCGTA----CATTTATGACATCATGTAGTCACGAACGATAA-ACAAAA 414
ISM25B1C4   ATATGTATATATATATGTGTGCATTTATGACATAATGTAGTCGGGAAGAAGAATACAAAA 420
            ******** * *          ****   **** ********       *  * ******

ISM335B1A2  ATGGGGCCGC-CAGGTAGTACTGGTATGCAGGAGGATCGTATTGATGAACGAAGTGCCAA 464
ISM335B1B6  ATGGGGCCGC-CAGGTAGTACTGGTATGCAGGAGGATCGTATTGATGAACGAAGTGCCAA 464
ISM335B1B7  ATGGGGCCGC-CAGGTAGTACTGGTATGCAGGAGGATCGTATTGATGAACGAAGTGCCAA 464
ISM25B1A11  ATGG---CGC-GAGGTCGTGGTGGTG---GGGATGGTATTGAGCATGATAAAGATGCCAA 463
ISM25B1C9   ATGG---CGC-CAAGCAGTGCCGGTACGCAGGAGGATCCTATTGATGAACGAAGTGCCAA 470
ISM25B1C4   ATGGGGCCCCAGAAGCCGAGTC-------AGGAAGAGTATAATAAAGTCAACAATGCGAA 473
            ****   * *  * *  *            *** *    *    * *       *** **
```

```
ISM335B1A2  ACATTTATTGGATAGCATAGGGAAAAAAGTGCAT---GACAAAGTGGAAAAGGATGATGC 521
ISM335B1B6  ACATTTATTGGATAGCATAGGGAAAAAAGTGCAT---GACAAAGTGGAAAAGGATGATGC 521
ISM335B1B7  ACATTTATTGGATAGCATAGGGAAAAAAGTGCAT---GACAAAGTGGAAAAGGATGATGC 521
ISM25B1A11  ACATTTATTGGATAGCATAGGGGAAAAAGTGTACAGAGAAAAAGTACAAAGTGATGATGC 523
ISM25B1C9   ACATTTATTGGATAGCATAGGGAAAAGAGTGCAC---AAAGAAGTGAAAAAGGAAGCTGA 527
ISM25B1C4   GGATCTTTTCGATTTAATTGGAAAATATATAGAA---AAAAAAGTGC---GTGATGCTGC 527
              ** * ** ***   ** **  **    *  *     *  ****       ** * **

ISM335B1A2  TAAAAATTATATTGATGATTTGAAAGGAGATTTGGCAAGCGGAACAAGTTCTCA----TT 577
ISM335B1B6  TAAAAATTATATTGATGATTTGAAAGGAGATTTGGCAAGCGGAACAAGTTCTCA----TT 577
ISM335B1B7  TAAAAATTATATTGATGATTTGAAAGGAGATTTGGCAGGCGGAACAAGTTCTCA----TT 577
ISM25B1A11  TAAAAATTATATTGGTGAATTGAAAGGAGATTTGAACAAAGCAACAAATCGTAG----TT 579
ISM25B1C9   ACAACGTAGTAGGAGTGATTTGAAAGGAAGTTTGTCATTTGCAACATTTTCTGT----TG 583
ISM25B1C4   TCTAGAACGTAAGGGAAATTTGAAAGGAAATTTAAAAAGCGCAAAATATAGAGAAGGCTA 587
               *     **      * *********  ***       * ** *  *         *

ISM335B1A2  CGGAATTAGTTA-GCAC-CGATAAAACATGCAAACTTGTAGATGATTATTATAATGAGCG 635
ISM335B1B6  CGGAATTAGTTA-GCAC-CGATAAAACATGCAAACTTGTAGATGATTATTATAATGAGCG 635
ISM335B1B7  CGGAATTAGTTA-GCAC-CGATAAAACATGCAAACTTGTAGATGATTATTATAATGAGCG 635
ISM25B1A11  CGGAATTAAGTATGCAG-CCTTGATCCGTGCATGTTTG---------ATTATACTGCACG 629
ISM25B1C9   TGGAATCAG-CATACAC-CACAGATCCGTGCCAACTTATAAAAGATAAAGGTCATAAACT 641
ISM25B1C4   TATCATCGAGCATGCAAACACAAATATATGTCATTTAATACATACACATGATACAAATGT 647
              **     *  **  *    *    **     *           *   *

ISM335B1A2  TGTTAATGGTGGTGGTGAACGGCATCCGTGCGTAAATGGAACA---GTAGAATATGTAAA 692
ISM335B1B6  TGTTAATGGTGGTGGTGAACGGCATCCGTGCGTAAATGGAACA---GTAGAATATGTAAA 692
ISM335B1B7  TGTTAATGGTGGTGGTGAACGGCATCCGTGCGTAAATGGAACA---GTAGAATATGTAAA 692
ISM25B1A11  TCTTGGTACTAACAGTAACAGGTATCCGTGCGCTAATAGATCA---CCAG---------T 677
ISM25B1C9   TCTTGGTGCTCGCGGTGA------TCCGTGCAAAAAAGACACAAACGGAAACAATGTAGA 695
ISM25B1C4   TACTGAGGGGCATGGAAAAGAGTATCCTTGTGCAAATAGATCA----GATA--------T 695
            *  *          *  *      *** **    **     **     *

ISM335B1A2  CCGTTTTTCGGATACACTTGGTGGCCAATGCACTGATCATAGAATAAAAGGTAATGAACG 752
ISM335B1B6  CCGTTTTTCGGATACACTTGGTGGCCAATGCACTGATCATAGAATAAAAGGTAATGAACG 752
ISM335B1B7  CCGTTTTTCGGATACACTTGGTGGCCAATGCACTGATCATAGAATAAAAGGTAATGAACG 752
ISM25B1A11  TCGTTTTTCCGATGAAAGCCGAAGCCAATGTACACAAAATAGAATAAAAGATAGTAC--- 734
ISM25B1C9   CCGTTTTTCGGATAAACAACAAGCAGAATATGATAACAAAAAAATAAAATGTAGTA---- 751
ISM25B1C4   TCGTTTTTCTGATAAACAAGGAGCAGAATGTGATGAGAAAAAAATAAGAGATAATGAA-- 753
             ******** ***  *          ***      *  * * ***** *  ** *

ISM335B1A2  TAATAAAACTGGTGGAGCATGTGCTCCACTCAGACGATTACATTTATGTGACAAAAATAT 812
ISM335B1B6  TAATAAAACTGGTGGAGCATGTGCTCCACTCAGACGATTACATTTATGTGACAAAAATAT 812
ISM335B1B7  TAATAAAACTGGTGGAGCATGTGCTCCACTCAGACGATTACATTTATGTGACAAAAATAT 812
ISM25B1A11  TAGCGGTACTGTAGGAGCATGTGCGCCTTTTAGACGATTATCTGTATGTGATTATAATTT 794
ISM25B1C9   -----ATAGTGAAGGAGCTTGCGCGCCGTTCAGACGATTACATTTATGCAACAAAAATAT 806
ISM25B1C4   -GACGACAGGGTAGGAGCATGTGCTCCATATAGACGATTACATCTATGCGACCAACATTT 812
                  *  *  ***** ** ** **    *********  * ****  *  *  ** *

ISM335B1A2  GGAAAAAATGGACGCAAATAATTATGATAGTGGTAAAGCTACGCATACGTTGCTCTCCGA 872
ISM335B1B6  GGAAAAAATGGACGCAAATAATTATGATAGTGGTAAAGCTACGCATACGTTGCTCTCCGA 872
ISM335B1B7  GGAAAAAATGGACGCAAATAATTATGATAGTGGTAAAGCTACGCATACGTTGCTCTCCGA 872
ISM25B1A11  AGAAAAAATAAGCACTAAAAAA---------AACAAAGCTAGACATAAGTTGTTGTTAGA 845
ISM25B1C9   GGTAAAAATGGACACAAATAATGATGATAGTAGTAAAGCTAAACATAATTTATTGTTAGA 866
ISM25B1C4   ATCGCACATGAAAGCTGAAAAA---------ATTAATACTAAAGATAATTTGTTGTTAGA 863
                 * **     *  * **             **  ***   ***  **  * *  **

ISM335B1A2  GGTGTGTCTTGCAGCAAAATATGAAGGAGAATCAATAAAAGATGATCATGCGCAATATCA 932
ISM335B1B6  GGTGTGTCTTGCAGCAAAATATGAAGGAGAATCAATAAAAGATGATCATGCGCAATATCA 932
ISM335B1B7  GGTGTGTCTTGCAGCAAAATATGAAGGAGAATCAATAAAAGATGATCATGCGCAATATCA 932
ISM25B1A11  GGTGTGTATGGCAGCAAAATACGAGGCAGAGTCACTACAAGGTTATTATGGTATATATGA 905
ISM25B1C9   CGTGTGTATGGCAGCAAATTACGAGGCAGAGTCATTAATAACTTATCATGATCAACATCA 926
ISM25B1C4   AGTGTGTCTTGCAGCACAATATGAAGGACAATCAATAAGAGTTGATCATGATAAATATAA 923
             ****** * ******* * ** ** * * * *** **  *  * ** ***    * ** *

ISM335B1A2  AGCAAAATATAATGATTTCCGTACCAATATATGTACTGAGTTAGCACGAAGTTTTGCAGA 992
ISM335B1B6  AGCAAAATATAATGATTTCCGTACCAATATATGTACTGAGTTAGCACGAAGTTTTGCAGA 992
ISM335B1B7  AGCAAAATATAATGATTTCCGTACCAATATATGTACTGAGTTAGCACGAAGTTTTGCAGA 992
ISM25B1A11  TGCAAAATATCACGATACTGGTTTTACAATATGTACTGCATTAGCACGAAGTTTTGCAGA 965
ISM25B1C9   AATGAC------TAATGTGGGTTCTCAATTATGTACCGAGTTGGCACGAAGTTTTGCCGA 980
ISM25B1C4   ATTAGACAATGATAATTCTGGTTCTAAATTGTGTACTGAGTTAGCACGAAGTTTTGCTGA 983
                      **    **       * ***** *  ** ************** **
```

```
ISM335B1A2  TATAGGAGATATTGTAAGAGGAAGAGATCTGTATCTGGGTTATGATCAAAAA-------- 1044
ISM335B1B6  TATAGGAGATATTGTAAGAGGAAGAGATCTGTATCTGGGTTATGATCAAAAA-------- 1044
ISM335B1B7  TATAGGAGATATTGTAAGAGGAAGAGATCTGTATCTGGGTTATGATCAAAAA-------- 1044
ISM25B1A11  TATAGGAGATATTGTAAGAGGAAGAGATCTGTATCTTGGT---AATCCAGAA-------- 1014
ISM25B1C9   TATAGGTGACATTATACGAGGAAAAGATCTATATCTTGGCAATAAAAAAAAAAAGCTAAA 1040
ISM25B1C4   TATAGGAGACATTATACGAGGAAAAGATCTGTATCTCGGT---AATCCGCAA-------- 1032
            ****** ** *** ** ****** ****** ***** **     *     **

ISM335B1A2  --GAAAAAGACCG--AAGAGAAAATTTAGAAAAGAATTTGAAAGAAATTTTCAAGAAAAT 1100
ISM335B1B6  --GAAAAAGACCG--AAGAGAAAATTTAGAAAAGAATTTGAAAGAAATTTTCAAGAAAAT 1100
ISM335B1B7  --GAAAAAGACCG--AAGAGAAAATTTAGAAAAGAATTTGAAAGAAATTTTCAAGAAAAT 1100
ISM25B1A11  --GAAATAAAACA--AAGACAACAATTAGATGAGAATTTAAAAACGATTTTTAAGAATAT 1070
ISM25B1C9   TGGAAAAGAAACAGAAAGGGATCAATTAGAAAGTAAGTTGAAAGAAATTTTCGGGGATAT 1100
ISM25B1C4   --GAAAGTGCACA--AAGAAAACAATTAGAAAAGAATTTGAAAGAAATTTTCAAGGAAAT 1088
              ****     *   ***  *  * *****    ** ** ***   *****   * * **

ISM335B1A2  ACATGGAGAATTG---AAGGATCCAAA------------AAAATCTCATTATAATGATCC 1145
ISM335B1B6  ACATGGAGAATTG---AAGGATCCAAA------------AAAATCTCATTATAATGATCC 1145
ISM335B1B7  ACATGGAGAATTG---AAGGATCCAAA------------AAAATCTCATTATAATGATCC 1145
ISM25B1A11  ATATGAGAAATTATTAGAGGATAACAAGACGAATGGTGTAAAAGACCGCTACGAAGATAA 1130
ISM25B1C9   ATATAAGGACGTGACGAGAGGGAAGAAGGAGGAG---ATAGAAAGGCGCTACGGAAG--- 1154
ISM25B1C4   ACATAGTGGATTGACGACGACGAACGG---------CGCACAAGCTCGCTACGGAGG--- 1136
            * **       *                          * **   *  **

ISM335B1A2  TGATGGAAATTTTTATCAATTAAGAGAAGATTGGTGGACTGCGAATCGGCACACCGTGTG 1205
ISM335B1B6  TGATGGAAATTTTTATCAATTAAGAGAAGATTGGTGGACTGCGAATCGGCACACCGTGTG 1205
ISM335B1B7  TGATGGAAATTTTTATCAATTAAGAGAAGATTGGTGGACTGCGAATCGGCACACCGTGTG 1205
ISM25B1A11  TGACGGAAATTATTATAAATTAAGGGAAGATTGGTGGACTGCGAATAGGCACACCGTGTG 1190
ISM25B1C9   TGATGGAAATTATTATAAATTAAGAGAAGATTGGTGGACGGCGAATCGAGAAACAGTATG 1214
ISM25B1C4   TGATGAAAATTATTTTCAATTACGAGAAGATTGGTGGAACGCTAATAGACAAGAAGTATG 1196
            *** * ***** ** * ***** * ************  ** *** *  *    ** **

ISM335B1A2  GGAAGCAATTACATGTAGCAAGAAGCTAGCAAATACTCATTATTTTCGACAAACGTGCAA 1265
ISM335B1B6  GGAAGCAATTACATGTAGCAAGAAGCTAGCAAATACTCATTATTTTCGACAAACGTGCAA 1265
ISM335B1B7  GGAAGCAATTACATGTAGCAAGAAGCTAGCAAATACTCATTATTTTCGACAAACGTGCAA 1265
ISM25B1A11  GAAAGCACTAACATGTGACAACAGGCTAGGGGGTTATTCATATTTTCGACAAACGTGCAA 1250
ISM25B1C9   GAAAGCTATCACATGTAA-GGCGGAC--GCAAGTAGTGCATACTTTCGAGCAACGTGCGA 1271
ISM25B1C4   GAAAGCAATTACATGTGATGCTGGG------AATGCTCAATATGTTGGACTTACATGTTC 1250
            * ****  * ******                 *  *   **  ** **   ** **

ISM335B1A2  TGGAGGAGAACAAACTAAAGGTTA-CTGCCGAT-GTGACGACAAGCCAAAGGCTGGCAAC 1323
ISM335B1B6  TGGAGGAGAACAAACTAAAGGTTA-CTGCCGAT-GTGACGACAAGCCAAAGGCTGGCAAC 1323
ISM335B1B7  TGGAGGAGAACAAACTAAAGGTTA-CTGCCGAT-GTGACGACAAGCCAAAGGCTGGCAAC 1323
ISM25B1A11  TGGAAAAGAACCAACTAAAGGTTA-CTGCCGGT-GTAACGGCGACC---AGCCAGGTAAG 1305
ISM25B1C9   TAGTGCTGATAAAA--AAGGTCCATCTGTAGCTAGAAACCAATGCCGGTGTGACGGCGTA 1329
ISM25B1C4   TGAGG--GAAGAAG------------TGCGACTCATGAAAAATGCACATGCGCTAGTGGA 1296
            *      **   *             **    *    *                  *

ISM335B1A2  GGCGACGTAAATATTGTCC---CCACA---TATTTTGACTACGTCCCTCAGTTTCTTCGC 1377
ISM335B1B6  GGCGACGTAAATATTGTCC---CCACA---TATTTTGACTACGTCCCTCAGTTTCTTCGC 1377
ISM335B1B7  GGCGACGTAAATATTGTCC---CCACA---TATTTTGACTACGTCCCTCAGTTTCTTCGC 1377
ISM25B1A11  GACAATCCAAATACCGATC---CCCCAACCTATTTTGACTACGTGCCGCAGTATCTTCGC 1362
ISM25B1C9   AAGAGCGCAAATGCCGACCAGGTCCCCACATATTTTGATTATGTGCCGCAGTATCTTCGC 1389
ISM25B1C4   GA------------TGTTC---CTACA---TATTTTGACTACGTGCCACAGTTTCTTCGC 1338
                        *  *      *    ******** ** ** ** **** *******

ISM335B1A2  TGGTTCGAGGAATGGGCCGAAGACTTA 1404
ISM335B1B6  TGGTTCGAGGAATGGGCCGAAGACTTA 1404
ISM335B1B7  TGGTTCGAGGAATGGGCCGAAGACTTA 1404
ISM25B1A11  TGGTTCGAGGAATGGGCCGAAGACTTA 1389
ISM25B1C9   TGGTTCGAGGAATGGGCCGAAGACTTA 1416
ISM25B1C4   TGGTTCGAGGAATGGGCCGAAGATTTA 1365
            ************************ ***
```

Curriculum vitae

Biography

Name:Joseph Paschal MUGASA
Date of Birth: 13th February 1972
Place of birth: Morogoro, Tanzania
Nationality: Tanzanian
Languages: English and Swahili

ADDRESS:

Ifakara Health Research and Development Centre (IHRDC)
P.O. Box 53 Ifakara, Tanzania.
Tel +255 23 2625164
Fax +255 23 2625312
Mobile + 255 784 412940
E-mail: *jpmugasa@ihrdc.or.tz , Joseph.Mugasa@unibas.ch*

EMPLOYMENT

2003-present Research Scientist, Ifakara Health Research and Development Centre (IHRDC)

1999-2003 Research Assistant, Sokoine University of Agriculture, Department of Veterinary Microbiology and Parasitology

ACADEMIC QUALIFICATIONS

Sept. 2005-April 2008: PhD in Microbiology - University of Basel, Switzerland and IHRDC, Tanzania. Thesis: "**Expression of *Plasmodium falciparum var* genes in naturally infected children from Tanzania".** Supervisor: Prof. Hans-Peter Beck

Sept. 2000-Dec. 2002: Master of Science in molecular epidemiology of human diseases and genetics. Katholieke Universiteit Leuven, Belgium. Thesis: "**Molecular characterization of *PapG*, the functional adhesins in Avian Pathogenic *E. coli* (APEC)".** Supervisor: Prof. Bruno Goddeeris

Sept. 1994-Sept 1998*:* Bachelor of Animal Science, Sokoine University of Agriculture (SUA), Tanzania

PROFESSIONAL DUTIES, WORKSHOPS AND MEETINGS

Jan. 2003-present. Involved in studying differential gene expression of *PfEMP1* in field isolates from children with severe malaria in Tanzania and molecular genotyping of anti-malarial drug resistance markers.

March 17- 28, 2007. World Health Organization special program for research and training in tropical diseases (WHO/TDR) and the national Centre for Genetic Engineering and Biotechology in Thailand (BIOTEC) -organized International training workshop entitled: *Functional Genomics of Malaria Parasites* Practical Course, Bangkok, Thailand**.**

March 13-14, 2007. *Computational Biology Congress*, Basel, Switzerland

Feb. 26-March 2, 2007. Advanced training course on bioinformatics, entitled: *Working with Pathogen Genomes*, Welcome Trust Genome Campus, Hinxton, Cambridge, UK

Nov. 31 – Dec. 8, 2004. Third Wellcome Trust/EMBO Workshop entitled: *HIV/AIDS and TB: the Way Ahead*, Cape Town, South Africa.

Nov. 1-5, 2004. African Malaria Network Trust (AMANET) Workshop entitled: *Molecular Biology and Immunology in Malaria Vaccine Development*, Witwatersrand University, Johannesburg, South Africa.

Sept. 15-17, 2003. *Medical Genetics Training Course for Developing Countries*, International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy

May 18 - June 6, 2003. East African Training Workshop*: Yeast Techniques in Malaria Research* and *Research in Protozoan Pathogens and Research Ethics*, Tanga, Tanzania. Funded by Seattle Biomedical Research Institute (SBRI) and National Institute for Medical Research (NIMR)

Sept 17. – Dec 17 2003: Attended training in *Molecular Biological Methods in Malaria Research*, Department of Parasitology and Infection Biology, Swiss Tropical Institute (STI).

Laboratory Experiences

Molecular Biology:

General molecular biology technologies, cloning and sequencing, quantitative real-time PCR, Magnetic beads isolation of specific mRNA, Transfection techniques, Reverse Transcription PCR, RFLP analysis

Microbiology:

Basic techniques in microbiology, *i.e* Microbial isolations and culture

Bioinformatics

Application of various programs for sequencing analysis and database search online ie DNAMAN, DANstar, Bioedit, MUSCLE, Mega, Arlequin, Phylip, NCBI, PlasmoDB,geneDB, Artemis and Artemis Comparison Tool (ACT), VectorBase

Biostatics

Basic Knowledge in statistical software (Stata, SPSS and Epi Info)

Awards

September 1999: Among winners of VLIR scholarships for Masters studying in Belgians universities

PUBLICATIONS

Mugasa JP, Qi W, Rusch S, Rottman M and Beck HP. Genetic Diversity of Expressed *Plasmodium falciparum var* genes from Tanzanian Children with Severe Malaria. Submitted to BMC

Rottmann M., Lavstsen T., **Mugasa J.P.**, Kaestli M., Jensen A.T., Muller D, Theander T., Beck, H.P. 2006. Differential expression of *var* gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. *Infection and Immunity*. 74(7): 3904-11.

Vandemaele F.J., **Mugasa J.P.,** Vandekerchove D., Goddeeris B.M. 2003. Predominance of the *papGII* allele with high sequence homology to that of human isolates among avian pathogenic *Escherichia coli* (APEC). *Veterinary Microbiology* 97(3-4): 245-57.

During my studies I attended lectures and courses of the following lecturers

N. Weiss, G. Pluschke, T. Smith, P.Vounatsu, HP Beck, I. Felger, R. Brun, Pieters, J, Cornelis, G Schwede T and Tanner M.

REFEREES

Dr Hassan Mshinda, PhD
Ifakara Health Research and Development Centre
P.O BOX 53, Ifakara, **TANZANIA**
Tel +255 232625164, Fax +255 232625312
E-mail: **hmshinda@ihrdc.or.tz**

Prof Marcel Tanner
Swiss Tropical Institute
Socinstrasse 57, CH 4002 Basel
Tel: +41-61-284 8287, Fax: +41-61-271 8654
E-mail: **marcel.tanner@unibas.ch**

Prof Hans-Peter Beck, PD, Ph.D
Swiss Tropical Institute
Socinstrasse 57, CH 4002 Basel
Tel: +41-61-284 8116, Fax: +41-61-271 8654
E-mail: **hans-peter.beck@unibas.ch**